THE EFFECT OF EXPERIENCE ON THE USE OF MULTIMODAL DISPLAYS IN A MULTITASKING INTERACTION

A Dissertation
Presented to
The Academic Faculty

by

Thomas Matthew Gable

In Partial Fulfillment
of the Requirements for the Degree
Psychology in the
School of Psychology

Georgia Institute of Technology
May 2018

THE EFFECT OF EXPERIENCE ON THE USE OF MULTIMODAL DISPLAYS IN A MULTITASKING INTERACTIO

Approved by:

Dr. Bruce Walker, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Mark Wheeler,
School of Psychology
Georgia Institute of Technology

Dr. Richard Catrambone,
School of Psychology
Georgia Institute of Technology

Dr. Carryl Baldwin,
Department of Psychology
George Mason University

Dr. Jamie Gorman,
School of Psychology
Georgia Institute of Technology

Date Approved: December, 14, 2017

To my parents and my wife for their never ending support and encouragement to achieve to the best of my abilities.

ACKNOWLEDGEMENTS

I would first thank my advisor, Dr. Bruce Walker for all that he has done to help me move through the dissertation as well as the graduate school process successfully and sanely. Without his decision to invite me to be a researcher in his lab and his guidance since then I would not be where I am today. I would also like to thank my committee members for all of their great feedback in the planning and writing process to help improve this research and document. Finally, I would like to thank my lab-mates for everything they do to support myself and each other in our research and out lives.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Theories and previous work suggest that performance while multitasking can benefit from the use of displays that employ multiple modalities. Studies often show benefits of these multimodal displays but not to the extent that theories of multimodal task-sharing might suggest. However, it is often the case that the studies investigating this effect give users at least one type of display that they are not accustomed to, often an auditory display, and compare their performance on these novel displays to a visual display, with which most people are familiar. This leaves a question open regarding the effects of longer-term experience with these multimodal displays. The current study investigated the effect of practice with multimodal displays, comparing two multimodal displays to a standard visuals-only display. Over the course of four sessions, participants practiced a list-searching secondary task on one of three display types (two auditory plus visual displays, and one visual-only display) while performing a visual-manual task. Measures of search-task and primary task performance along with workload, visual behaviors, and perceived performance were collected. Results of the study support previous work with regard to more visual time on the primary task for those using multimodal displays, and show that perceived helpfulness increased over time for those using the multimodal displays. However, the results also point to practice effects taking place almost equally across the conditions, which suggest that initial task-sharing behaviors seen with well-designed multimodal displays may not benefit as much from practice as hypothesized, or may require additional time to take hold. The results of the research are discussed regarding their use in research and applying multimodal displays in the real world as well as in how these results fit with theories of multimodal task-sharing.

CHAPTER 1: INTRODUCTION

We as humans commonly perform more than a single task at once. Sometimes this multitasking can easily be done at the same time, such as talking while walking. In other situations, these tasks are not as advisable to be done simultaneously such as texting and driving. In an effort to decrease the negative effect of performing these secondary tasks, some researchers and developers have employed multimodal displays to create better displays to be used in these dual tasks. While theory and research has suggested that the use of these sorts of multimodal displays can be effective in helping dual task users, the displays do not always result in the performance improvement expected when compared to a visuals-only interface. Assuming that the theories behind the use of these multimodal displays are true, it may be that participants are simply not practiced enough in using multimodal displays. By giving participants time to practice and become accustomed to employing other modalities to complete these types of tasks, research might reveal differences in performance and user strategies with these types of displays not seen before. In addition, by watching how participants change the way they interact with these types of displays as they gain practice we may gain a better understanding about the process of learning to use multimodal displays. This document looks further in depth to explore the literature behind these areas of research and then describes a line of research that investigated the question at hand – How does increased experience with a multimodal display change performance in a multitasking situation?

1.1 Multitasking While Driving

A commonly undertaken – but unsafe – way that humans multitask is by driving and performing additional, non-driving related tasks. These “secondary tasks” such as texting (Drews, Yazdani, Godfrey, Cooper, & Strayer, 2009) or simply interacting with the car's entertainment (Stutts, Reinfurt, Staplin, & Rodgman, 2001) have been shown to be detrimental to safe driving behaviors. However, even with these potential hazards people continue to perform secondary tasks, as illustrated by the 26,000 police-reported crashes in 2010 alone that were linked to distraction due to the use of a device or in-vehicle controls (National Highway Traffic Safety Administration, 2013). The frequency and high potential hazard of participating in these tasks makes the area ripe for research as it means that the results can be relatable to the general population and media, it can be easier to find potentially experienced participants than in other dual task situations, and the results can have a greater impact when applied directly to the population. For these reasons, multitasking while driving is a heavily researched area of dual-task interaction (Chisholm, Caird, & Lockhart, 2008; Harvey, & Carden, 2009; Lasch, & Kujala, 2012).

1.1.1 List Searching

One common type of in-vehicle dual tasks is searching a list, such as browsing a song list on a music player or smartphone, or interacting with a menu system in a car infotainment display to find a song, contact, or other information the driver cannot remember explicitly. In these situations, the user applies recognition memory instead of recall, meaning that speech recognition devices cannot be used. These types of interactions can happen while using both brought-in and built-in devices, and are most often done via a visual display. Previous work has found that using such a music player device negatively affects driving performance and increases driver anxiety, compared to not using such a device (Harvey & Carden, 2009), and there is evidence of increases in fixation time on the secondary task and perception response time (PRT) when performing a difficult music selection task (Chisholm, Caird, & Lockhart, 2008). These negative effects on driving and the visual fixations of participants also do not appear to significantly decrease with practice when using visual displays (Chisholm et al, 2008), suggesting people are already fairly experienced with theses types of tasks.

1.2 Applying Multimodal Displays

Some researchers approach the design of dual task systems when visual demand is high for one task, such as in driving, through the use of other modalities to display the information in one of the tasks. Nees and Walker (2011) recommend the use of auditory displays in situations where visual demand is high on the secondary task within the vehicle cockpit, such as a list-searching task. They suggest that visual displays in the vehicle have a fundamental shortcoming: they require high levels of visual attention when ideally the driver should have visual attention focused on the driving task. This is particularly the case for more complicated, interactive displays as opposed to “on-off” displays, with basic coding schemes and no interaction. This differentiation is important to point out, as a simple “on-off” display codes very little information into the display, such as a simple beep or buzz to let the individual know that there is a change in some status. These basic displays also do not necessarily require any sort of direct interaction with the interface from the user, nor do they code any complicated information into the display. Whereas basic displays are interesting to research and have their purposes, the aim of the current document is focused on more interactive, sophisticated displays.

1.2.1 Multiple Resources Theory

A theory describing the use of these alternative modalities is Multiple Resources Theory (MRT). The theory describes the allocation of mental resources to complete multiple tasks at once, stating that if the tasks are done in different modalities it will lead to better performance of the two tasks than doing them both in one modality (Wickens, 2002; Wickens, 2008). In other words, MRT states that if two tasks are completed at the same time but distribute mental processing (aka workload) across multiple human perceptual pathways (aka modalities) such as vision and hearing instead of both being in one modality such as vision only, it can facilitate increased performance (time-sharing). This means that by having a task performed primarily via one modality (i.e., the visual modality of the driving task) and another task performed primarily with another modality (i.e., using the auditory modality for list searching) an individual can have overall better performance on the two tasks than if they attempted to perform both tasks visually. It should be noted here that MRT also states that if there are not enough resources available, then there will be a decrease in performance, with performance either decreasing on both tasks, or more likely, one of the tasks being released to ensure the other one is completed.

Research to investigate these claims of multimodal time-sharing for interactive displays seems to support the basis of the theory. In one study, lower dwell time off the primary driving task (i.e., more eyes on the road) was found when participants used an interactive multimodal interface as compared to a visuals-only interface (Chisholm et al., 2008). In another study an interactive multimodal interface led to better performance on hazard detection, navigation and driving tasks, and lower workload (Liu, 2001). These results suggest that more visual time and mental workload was spent on the primary task when using the multimodal displays and that it decreased workload for the dual task scenario as a whole. A similar study found that two novel auditory interfaces were preferred by drivers over the visuals-only interface, and that participants had better driving performance, and lower perceived workload, although task completion times were slower for longer tasks (Sodnik, Dicke, Tomažič, & Billinghurst, 2008).

Similar increased times to completion were found in another study when a visual display was compared to a multimodal display while driving; however, the multimodal display again decreased the risk created by interacting with the interface (Zhao et al., 2013). Although these results are promising for auditory or multimodal interfaces, the time it takes to complete an action is of particular concern when considering the adoption of these types of interfaces. This concern is displayed in research that found the decrease in completion time was an important enough factor to the point of users abandoning the auditory cues when the secondary task is higher priority, which some call willingness to engage (Brumby, Davies, Janssen, & Grace, 2011; Ranney, Mazzae, Garrott, & Goodman, 2000). This abandonment, caused by slower performance time or potential unfamiliarity and confidence with the auditory displays, may be traced back to two potential factors: 1) the auditory displays are too slow to be used as compared to the visual interfaces; or 2) such research often employs novel interaction methods that the multitaskers are not accustomed to using. Both of these factors create an issue of unfair comparison for the auditory and visual displays as the users are taking longer to get the information required and are only starting to get familiar with such multimodal interfaces by the time the study is completed.

1.2.2 Advanced Auditory Cues

To address the issue of slow auditory feedback, advanced speech-based auditory cues for list navigation were developed (Jeon & Walker, 2011; Walker, Nance, & Lindsay, 2006). One of these cues, called a Spearcon, is a brief sound produced by speeding up a spoken phrase, even to the point where the resulting sound may no longer be comprehensible as a spoken word (Walker, Nance, & Lindsay, 2006). These auditory cues can be very useful for short, well-known menus and have been shown to be better than earcons (abstract, non-speech auditory representations of items) in regards to rates of learning (Dingler, Lindsay, & Walker, 2008; Palladino, & Walker, 2007). Another type of advanced auditory cue is a Spindex (i.e., speech index), a set of short non-speech auditory cues based on the pronunciation of the first letter of each menu item (Jeon & Walker, 2011). Spindex cues are particularly useful in alphabetical lists as they allow for fast movement down a large list and are faster to learn than a Text-To-Speech (TTS) interface alone (Jeon & Walker, 2011). Both Spearcon and Spindex cues are usually used to enhance a typical auditory menu, which would otherwise consist only of simply spoken menu items (TTS), often being followed by TTS and made to be interruptible to allow for rapid movement through a list.

Research has shown these cues to be fairly helpful when applied to a search task. In one study participants performed a search task on a mobile phone by flicking, wheeling, or tapping while hearing different advanced auditory cues or only seeing the visuals. Results indicated that when participants heard the advanced auditory cues, they had significantly faster search times and lower subjective workloads as compared to no auditory cues using the same input methods (Jeon, Walker, & Srivastava, 2012).

These advanced auditory cues have also been applied to the driving context. In one such study, Jeon, Davison, Nees, Wilson, and Walker (2009) found decreased subjective cognitive workload and item selection time, as well as preferences for the auditory system when participants completed the search task on a head unit (i.e., car stereo or entertainment interface) while also performing a driving like-task called a ball drop game. Some analogous results were found in a similar study done on a mid-fidelity simulator and head unit (Jeon et al., 2012). Gable, Walker, Moses, and Chitloor (2013) and Gable (2015) also investigated this space, using eye-tracking to determine how the auditory cues affected visual behaviors. Results in both studies showed increased dwell time toward the primary task in the list-searching conditions when Spindex cues were applied as compared to the visuals-only interaction but did not show the type of conclusive results that could be hypothesized by MRT for eye-tracking or other measures such as driving performance, search task performance, or workload. These results of more visual time on the road could have large impacts on real-world users as previous work has suggested that even a small amount of time that drivers have their eyes off of the road can be harmful. One such study found that if a driver's eyes were off the road for more than 2 seconds in the 5 seconds before a hazardous event, the potential for a crash or near crash doubled (Klauer, Dingus, Neale, Sudweeks, & Ramsey, 2006).

1.3 Experience and Practice

As seen above, some of these previous studies investigating multimodal displays show significant differences as compared to visuals-only displays, just as hypothesized by MRT. However, they have not all been as convincing as one might expect. One factor in much of the multimodal research that is often not discussed or focused on is the level of practice with the interface in a study. In most of the studies in the space, the novel multimodal displays are given to participants with very little practice, usually after some limited training. This leads to the question of how practice might affect the outcomes of research with multimodal displays and how that relates to MRT.

This could be a large issue as people are very practiced at interacting with visual interfaces, however they are often not as used to using auditory cues to complete these sort of tasks. It may be that participants in these studies are simply not being trained as highly as would be needed for participants to truly employ multimodal displays (such as advanced auditory cues) in the most effective way or for MRT to show maximal effects. In addition, watching how participants change the way they interact with these types of displays and complete the dual tasks may provide a better understanding about the process of learning to use multimodal displays and how to interpret and apply MRT more effectively.

1.3.1 Effects of Practice on Task Performance

Schneider (1985) lays out a number of considerations regarding the process of developing high performance skills. The first of these is that of the extended practice function: as practice increases, the rate of performance improvements decreases. This is often called the power law, or the "ubiquitous law of practice" (Newell & Rosenbloom, 1981) and is expected to occur in any complex skill acquisition process. The second factor to consider is that a task with more stable components will have a faster learning process than a task with varied components. This means that the more methodical and rule-based the reactions required by users, and the less variable reactions required, the faster the skill will be learned. Schneider then discusses the process of learners going through stages with different strategies to complete the task as they gain more experience, eliminating bad strategies, as they get better at the task. This means that as people gain experience with a task they are able to drop bad strategies in completing that task, and employ new and better processes for completing the tasks in their place. The last characteristic to be expected is that of learning to develop time-sharing abilities, the ability to complete the practiced task and additional tasks, splitting attention and workload across the multiple tasks as needed to compete both successfully.

These factors are important to consider when talking about practice of a skill, in this case the use of multimodal displays. First, the law of practice may allow users of novel multimodal displays to quickly get to a decent level of ability, whereas giving the same additional practice with a visual display to participants may not significantly increase their abilities more than they already have from their high level of previous knowledge. This suggests additional training in all conditions may be warranted. Second, the type of tasks that these secondary tasks often employ – in particular list searching – is very stable, which means it may be a quick skill to learn. Third, some of the results seen in other studies after just a bit of practice may have been with participants still using bad strategies with the novel multimodal displays. Considering Schneider's (1985) discussion of improving strategies as practice goes on, giving participants additional time with these displays may allow them to switch strategies multiple times, increasing their performance more than in other studies. However, it is important to consider in the current study that the goal was to look at performance in a dual-task situation, so it was deemed best for participants to practice in a dual task situation since their changing strategies for the task could differ if they practiced in a single task environment. This was particularly important for the use of the multimodal cues, as leaving participants to perform the search task alone may not have pushed them to employ the auditory cues and instead just built strategies around using visuals for all of the conditions. Finally, the time-sharing abilities factor is very important in a dual task situation such as this. Giving participants more time to be comfortable and increase time-sharing abilities will help to allow us to see the full potential for auditory displays in this setting.

1.3.2 The Factor of Automaticity

Another factor that is tied into the dual task literature, training, and time-sharing is automaticity. With automaticity the idea is that highly practiced tasks can be done at high performance levels without very much attention or workload on the user's part due to the low level of demands it places on extensively trained users (Lewandowsky & Thomas, 2009). An example of this would be the process of driving home from work, or tying your shoes, tasks that at one point took extensive thought and workload to complete but are now done very easily and without effort. Lewandowsky and Thomas (2009) broke down the main two theories for the process of developing automaticity in dual tasks. One of these processes is that people take what are originally independent declarative steps and combine them into more overarching rule-based models of knowledge (Speelman & Kirsner, 1997), resulting in the individual having fewer decisions to make in the completion of the task. The other approach is that of pure memorization (Logan, 1988), where the more processes of a task the individual memorizes the reactions to, the more automatic the task becomes. With either model, research has supported the idea that people can reach automaticity to a point where the task interferes less with other ongoing tasks, leading to parallel processing and therefore time-sharing. Increasing the automaticity of the search task in participants could reveal more trends unseen in other research.

1.3.3 Practice and Multimodal Displays

While practice and automaticity have not been extensively discussed in the literature regarding MRT and auditory displays, they are nevertheless important. If MRT is to be used as a supporting theory for using multimodal displays, the effect of practice and other real life issues must be considered to help inform what the actual time-sharing and performance levels will be. Although not often addressed in the MRT literature, this idea is discussed in one paper where Wickens says that performance on a multimodal task could be highly influenced by the skill level of the user and therefore the practice of that task (Wickens, 2002). This suggests that MRT makes the assumption of equal experience on all tasks being performed, and the modality they are being completed on, an assumption that some researchers have falsely made or ignored when comparing visual interfaces to novel multimodal displays.

The reason for this lack of equal experience lies in the fact that each day those in the Western world gain experience with technology such as their phones, computers, or other electronic devices. Most often that experience is done through the visual modality. This previous experience with visuals, and therefore potential levels of automaticity, may be one of the major limitations to research in the multimodal realm – almost everyone is an expert at visual interaction. This means that even if the visuals are different from what someone is used to, the interaction technique of visually looking at a device and then manipulating it physically is well practiced. Regarding visual-manual tasks, this suggests that people have had such extended practice doing visual-manual tasks that attempting to undergo a different type of task, such as an auditory-manual task, may be very different from previous experiences and therefore need more practice to reach similar levels of performance. If this is true, and people are substantially worse in these multimodal tasks due to lack of enough familiarization with the type of interaction then it may be that getting a higher level of familiarization with the multimodal auditory-manual task could allow people to better perform with these types of displays.

1.4 Current Research

The current research aimed to investigate the potential effects of extended experience on participant interaction with advanced auditory cue (AAC) displays as compared to a visual-only display. With previous work often showing similar or better performance to a visual-only display after only a small amount of experience, extended experience with these AAC displays may begin to show other differences in task sharing and task performance. The present research included two types of AAC multimodal displays (Spindex-TTS and Spearcon-TTS multimodal displays) as well as a control group of a visual-only display, all of which were randomly assigned.

The research had participants complete a secondary list-searching task while performing a primary visual-manual task. In this study the primary task was either a highly controlled “ball drop” game (Jeon, Davison, Nees, Wilson, & Walker, 2009; Wilson, 2016) that was calibrated to control for learning effects, or a driving task in a driving simulator. The ball drop game was used to confirm that participants built their new strategies, as they gained practice, for a dual task situation of performing the search task while doing a visual-manual task. This would ensure participants did not build strategies for the search task as a single task or focus on using visuals if they were in the multimodal conditions, allowing them to transfer skills more easily when dual-tasking in a driving simulation. The calibration of the ball drop game was also done to ensure that any changing behavior and performance on the secondary task was not driven by increased abilities in the ball drop task. The driving task was used as a realistic task to see how the learning effects could apply to the real world. Participants' performance on the search task as well as on the visual manual task (ball-drop or driving simulation) was collected repeatedly throughout the research. In addition, physiological and subjective measures of workload, visual behaviors, and perceived performance and preferences were collected. These measures were compared across training time and between display types to look for differences and interactions both within and between subjects.

The following two chapters describe a pilot (Chapter 2) and the full study (Chapter 3), which were done to investigate the potential learning effects of using the AAC displays over time and how these applied to real world use-cases. The pilot was done in an effort to determine the amount of training necessary for the participants in the full study to show learning effects. The full study then used this information to train participants with the displays and track this learning, but also compare performance on a real world use-case of driving at the beginning and end of the study.

CHAPTER 2: PILOT

A pilot study was run to ensure that participants in the full study had enough training time to display the effects of practice. In the pilot a group of six participants were trained over five one-hour training sessions (30 minutes of practice per session) to see where learning effects began to peak and where participants reached sufficient levels of training.

2.1 Research Questions and Hypotheses

RQ1.1. **How will practice affect the way participants perform the dual task?** This was studied by looking at the participants' performance on the tasks, their subjective and objective workload, perceived performance, and visual behaviors across the training blocks. The following hypotheses (H) were put forward to test.

H1.1.1 – Participants' secondary task performance would increase as they got more practice with the displays as seen through increased accuracy and decreased search times for the secondary task. This was expected due to the expected training effects, with practice increasing performance on tasks over time.

H1.1.2 – Workload would decrease as participants got more practice with the display over the blocks of training. This would be seen through decreased heart rate and lower NASA-TLX scores. This was hypothesized, as it should take participants less workload to complete the dual task as they received more practice and became more experienced.

H1.1.3 – Participants would report feeling more comfortable and have higher subjective performance with their assigned displays as they went through more training blocks as seen through high comfort and subjective performance scores on the session questionnaire. Again, this was expected due to the increased time with the displays and therefore expected increase in comfort.

H1.1.4 – Participants' visual behaviors would be less focused on the secondary task as practice with the displays increases, as measured through lower percent time eyes off the primary task, lower glance count rate, and lower average dwell length off the primary task. This was expected to occur as practice should allow participants to become accustomed to the display they are using and help them to form an approach for how to do their visual task sharing more efficiently.

RQ1.2. Will participants reach sufficient expertise and if so how long will it take? This was determined by looking for visual behaviors, workload, and the dual-task performance (accuracy and speed) to become less variable as practice increased. The participants in the visuals-only condition were used as the baseline for determining what this peak looks like as they are already experts at this type of interaction and should therefore be able to serve as a model of what the sufficient expertise looks like. To investigate this question the following hypotheses were put forward to test.

H1.2.1 – Changes in task performance, workload, visual behaviors, and perceived performance/comfort would begin to slow down as participants gain more practice with the displays. This was hypothesized due to the nature of learning being a curve and not linear and would be seen through the dependent measures previously discussed plateauing.

2.2 Methods

2.2.1 Participants

Six participants took part in the pilot. The participants had an average age of 25.17 (*SD*=2.32) and included 3 males and 3 females. They were recruited via word of mouth (see recruitment materials, Appendix A). All participants were required to have a valid driver's license in the United States for a minimum of 2 years; report normal or corrected to normal vision, hearing, and mobility; and avoid performing any strenuous exercise or caffeine intake for two hours prior to participating in each session. Participants received $10 of compensation for each hour that they participated.

2.2.2 Apparatus and Materials

List Search Displays. During the study 3 groups of 2 participants used different types of displays including a visuals-only condition (No-Sound), a visual + Spearcon + TTS (Spearcon) condition, and a visual + Spindex + TTS (Spindex) condition. Participants only used one of these displays throughout the entire study.

Visuals. The visuals used in all of the conditions were the same, with the names of the songs listed in alphabetical order and lines between each song name. A screenshot of this can be seen in Figure 2.1.

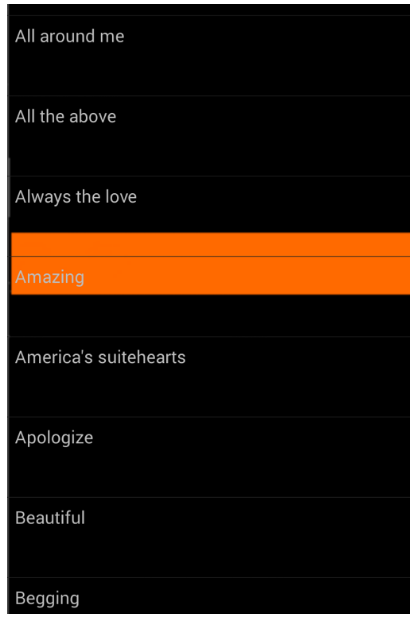


Figure 2.1 - Visual display used for all three conditions as displayed on the phone.

Text-to-speech cues. The TTS cues were generated using the AT&T Labs TTS Demo program with the male *Mike* voice (http://www.research.att.com/~ttsweb/tts/demo.php). These were then played as recorded from the website.

*Spearcon cues***.** The Spearcon cues were created by taking the TTS files for each song name and putting them through the Georgia Tech Sonification Lab's Spearcon generation algorithm. This algorithm compresses each TTS cue logarithmically but keeps its original sound frequency. The Spearcon displays were followed by the TTS cues after a 250-millisecond (MS) silent interval between the Spearcon and the beginning of the TTS. The cues were interruptible so that the auditory cues were overridden by the next cue if the participant moved down the list before the audio for the previous cue was finished playing.

Spindex cues. The Spindex cues used in the study were generated TTS for each letter of the list (e.g. “A”), totaling 26 sound files. The attenuated version of Spindex was used in this study as previous work has shown it to be the most preferred (Jeon & Walker, 2011). This meant that after the first item for each letter the following song names were announced with a 20dB decrease for the Spindex cue. As with the Spearcon cue the Spindex cues were followed by the TTS cue after a 250-MS silence.

List Search Task. The participants consistently performed the search task with or without auditory cues throughout each of the study blocks based on the condition they were randomly assigned to ahead of participant scheduling. The task was done on a Google Nexus One HTC1 Android smartphone running Android OS version 2.3.6. The 3.75-inch resistive touch screen displayed a list of 150 popular songs from 2009 and participants interacted with the list through kinetic flicking. While this interaction technique is not optimal for driving, it is a standard interaction mode for touch-screen devices (Lasch & Kujala, 2012). Each participant was allowed to choose the hand she wanted to use for the task before the study started, and then used that hand throughout the whole study. The participants placed their arm and hand on an armrest while their other hand was used for the visual-manual task.

Performance on the search task was measured through the time it took participants to find a correct song (in seconds), the number of songs searched, and the accuracy for each block. These data were calculated and stored by the phone and then uploaded following data collection to a computer.

Visual-Manual Task. During each block participants also performed the ball drop game. This was a visual manual vigilance task that consisted of seven vertical columns from which balls fell, and a paddle at the bottom that participants used to catch the balls. The participants controlled the paddle via the arrow keys on a standard keyboard and viewed the task on a 21.5” Dell LED monitor with resolution of 1920x1080. The goal was to catch as many balls as possible. The balls that were dropping from the top could start a maximum of two columns away from where the previous ball was released. This ensured a moderate but not extreme difficulty. The dropping rate of the game was also calibrated individually for each participant at the beginning of each session so that each participant caught between 75-85 percent of the balls that were released. This maintained difficulty level as participants gained practice. An image of this task can be seen in Figure 2.2.

Figure 2.2 - Participant completing the ball drop game while performing the song selection task.

Performance on the ball drop task was measured via accuracy (determined by the number of balls caught divided by the total number of balls released during each block) and the number of balls released. The accuracy measure represented the participants’ performance for each block while the measure of balls released per block represents the number of balls participants had the opportunity to catch during each block and are what is determined during the calibration setup at the beginning of each session due to the speed of which balls are released by the system. This data was calculated by the program and stored in a log file after each session.

Visual Behaviors. The visual behaviors of the participants were measured using a set FaceLAB 4 fixed eye trackers, which can also be seen in Figure 2 below the display. The system cameras were set to use either the dark iris or dark pupil method (depending on which method had higher tracking rates for that participant) to gather data at 60 Hz during the study. In the eye-tracking software a 3D model was created of the primary task screen to determine when the participant was looking at the ball drop task and when they were looking down at the secondary search task. Due to loss of tracking at certain angles, particularly when participants looked away from the primary task, any frames when the participant was not being tracked as being on the primary task was reported as off task.

The eye-tracking data were stored in log files and synced with the other data being collected. Eye-tracking values as defined by SAE J-2396 (Lamble, et al., 1999) were calculated to look at the visual behaviors including: percent dwell time off-task (calculated by dividing the number of frames where the participant was not on the primary task by the number of frames); mean glance frequency off-task per minute (the average number of glances off the primary task per minute); and mean glance duration off the primary task (the average period of time in ms that a participant had their eyes off the primary task).

Objective Cognitive Load. Workload was also measured physiologically through heart rate beats per minute (HR) and heart rate variability (HRV), which was collected throughout the study. The data was collected at 32 Hz with a NeXus-10 physiological monitoring and biofeedback platform. The system gathered EKG data at 256 Hz from the participants via a modified lead II configuration, meaning the ground under the left clavicle, the positive lead on the left lower ribs, and negative lead under the right clavicle. The physiological system collected and stored the data, which was matched via timestamp along with the primary task. The measures collected from the system included mean HR (BPM) in a block, and mean HRV per block.

Subjective Cognitive Load. Subjective workload in the pilot was measured through NASA-TLX on an additional computer using a mouse and keyboard after all blocks (Hart, 2006). The TLX measures six subscales of workload including effort, temporal demand, physical demand, frustration, performance, and mental demand. Data from the TLX survey was saved on a computer once completed. The data was measured on a 100-point scale for each of the dimensions and the single numerical value was output as a measure of total workload as well as each raw subscale to investigate the individual factors, known as the raw TLX (RTLX) (Hart, 2006).

Other Measures. Participants completed two types of surveys over the duration of the study: a demographics questionnaire (Appendix B) given during the first session of the study; and a preferences questionnaire (Appendix C) given for each training session regarding perceived performance on the cell phone and visual manual task for that session.

2.2.3 Procedure

At the beginning of the first session, participants were given the screening form

(Appendix D) to ensure they met the study criteria and then given the consent form (Appendix E) and asked to review it and sign if they agreed to participate. If they agreed to participate they reviewed the experiment instructions (Appendix F) and asked the experimenter any questions they had. Once the instructions were reviewed the experimenter showed the participants how to do the secondary search task, and introduced them to the display they would be using by letting participants use it for 1 minute, or until they said they were ready.

Following these introductory steps in Session 1 all other sessions in the pilot were the same. They began (or in Session 1, continued) after participants confirmed they met the criteria on the screening form and then beginning the process of setting up the physiological systems. The participants followed the heart rate instructions (Appendix G) to place the monitor pads on the their bodies while the experimenter was out of the room. The participants were then seated in front of the display and underwent the calibration process for the eye-trackers. The participants then performed the ball-drop task calibration (without the secondary task) to 75-85 percent accuracy. This accuracy level was chosen as it aimed to ensure performance was near participants’ highest abilities, but would not create a ceiling effect.

Once the systems were all set up the study blocks would begin. During these blocks the visual-manual task and cell phone search task were both started at the same time. The phone used TTS to announce which song the participant was supposed to search for and displayed the name on the screen. As the participant navigated the song list they received either visual or multimodal feedback to inform them of where they were in the list. When the participant believed she had found the song, she selected the song by pressing on the song name on the phone screen. Once a participant made a selection there was a 5 second break in the secondary task and then another song was announced. This timing was chosen to give participants a few seconds rest between each search task, and short enough to ensure that participants are consistently performing the dual task. Each block was 10 minutes long, with participants receiving a 5-minute break between each block to fill out the NASA-TLX and rest.

Each session in the pilot consisted of 3 blocks. After the first session participants filled out the demographics questionnaire (Appendix B) and then the preferences questionnaire (Appendix C), discussing their comfort and perceived performance on the tasks. In the other sessions, participants only filled out the preferences questionnaire. Finally, when the participants departed from the session they were given the payment for that session. In the final session the participants also received the debrief form (Appendix H) before they departed. The order of the blocks and events for the 5 sessions can be seen visually in Table 2.1.

Table 2.1 - Visual representation of the order of events for each session in the pilot. Note that the sizes of the cells are not representative to the actual time for each event in the session.

<table>
<tr><th>Session 1</th><th>Session 2</th><th>Session 3</th><th>Session 4</th><th>Session 5</th></tr>
<tr><td>Study intro</td><td rowspan="2">Screening and physiological setup</td><td rowspan="2">Screening and physiological setup</td><td rowspan="2">Screening and physiological setup</td><td rowspan="2">Screening and physiological setup</td></tr>
<tr><td>Screening and physiological setup</td></tr>
<tr><td>Ball drop setup</td><td>Ball drop setup</td><td>Ball drop setup</td><td>Ball drop setup</td><td>Ball drop setup</td></tr>
<tr><td>10-min ball drop</td><td>10-min ball drop</td><td>10-min ball drop</td><td>10-min ball drop</td><td>10-min ball drop</td></tr>
</table>

TLX	TLX	TLX	TLX	TLX
10-min ball drop	10-min ball drop	10-min ball drop	10-min ball drop	10-min ball drop
TLX	TLX	TLX	TLX	TLX
10-min ball drop	10-min ball drop	10-min ball drop	10-min ball drop	10-min ball drop
TLX	TLX	TLX	TLX	TLX
Questionnaire	Questionnaire	Questionnaire	Questionnaire	Questionnaire

2.2.4 Design

In the pilot the independent variables were the amount of training participants received and the three conditions participants were randomly placed in. The dependent variables included the list search performance (songs searched, accuracy, and time to find a song), ball drop performance (accuracy and balls caught), visual behaviors (percent time eyes off the primary task, glance rate off task per minute, and average dwell time off the task), heart rate measures (mean HR and HRV), NASA-TLX scores, and reported levels of annoyance, preferences, and perceived performance. No outliers were removed in the pilot due to the low number of data points and therefore unreliable determining of outliers. Analyses were also not performed on the data due to the low number of participants but instead the trends were used to determine learning rates.

2.3 Pilot Results

The results of the pilot data collection are included in the section below. The discussion of the collected data focuses on the data as a whole and less so between conditions as the between differences was focused on in the full study in Chapter 3. The descriptive data for the measures collected can be found in Appendix I.

2.3.1 List Search Performance

The descriptive data for measures of list search task performance including number of trials, percent accuracy, and average time to find a correct song can be seen in Table 1 in Appendix I.

Number of Searches. As seen in Figure 2.3 and in the descriptive data the number of searches per block trends upwards throughout the study, increasing by about 3 more searches on average across the 15 blocks. No real plateau was seen in the data collected, suggesting that participants may have continued to improve upon the number of searches per block at the same rate if allowed to continue training.

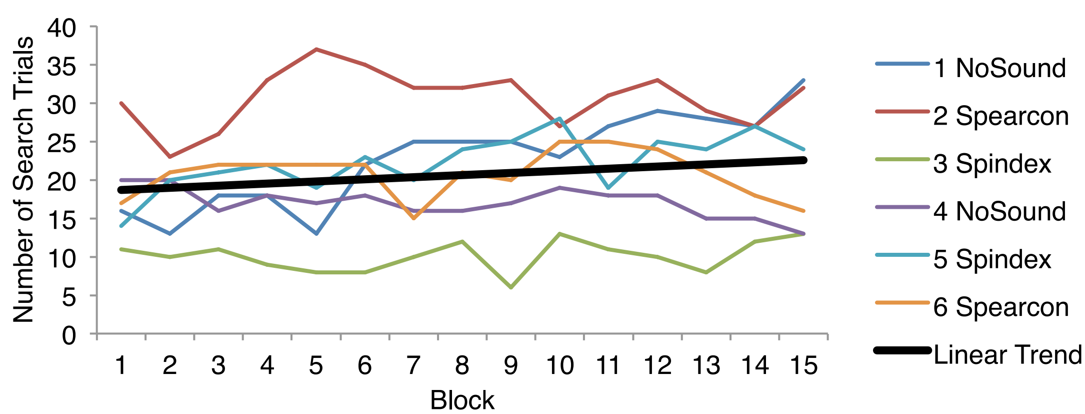


Figure 2.3 - Graph of the number of songs searched for each participant across the 15 blocks. The thicker black line shows a linear trend line for the 6 participants.

Percent Correct. Figure 2.4 shows the data for percent correct song selections over the 15 blocks. The trend seen in the graph and the data in Table I.1 show slight increases in percent correct, but decreasing standard deviation towards the end of the 15 sessions. This suggests that participants' learning effects may have been slowing down in the latter blocks and a potential plateau as they converged near 90 and above correct.

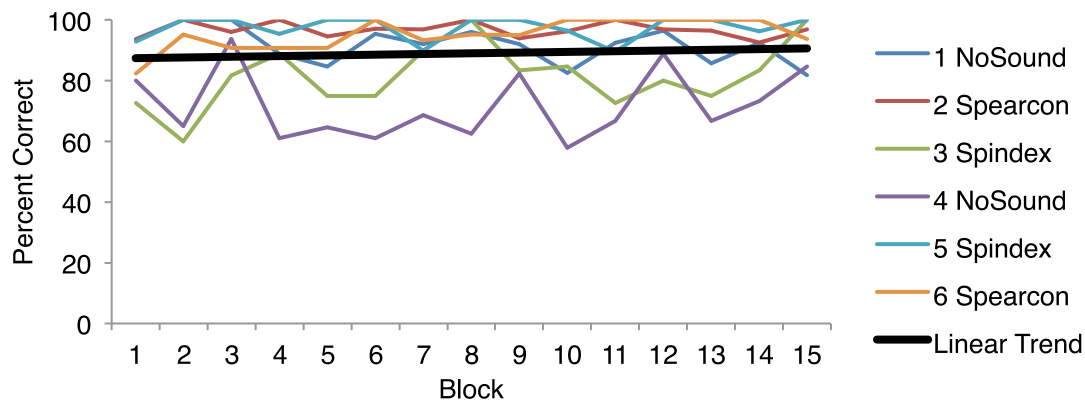


Figure 2.4 - Graph of the percent correct songs selected for each participant across the 15 blocks. The thicker black line shows a linear trend line for the 6 participants.

Average Correct Selection Time. As seen in Figure 2.5, the time to find a correct song was decreasing as would be expected through increasing number of searches and accuracy. While it seems no plateau had been reached as of yet it still shows that learning did occur.

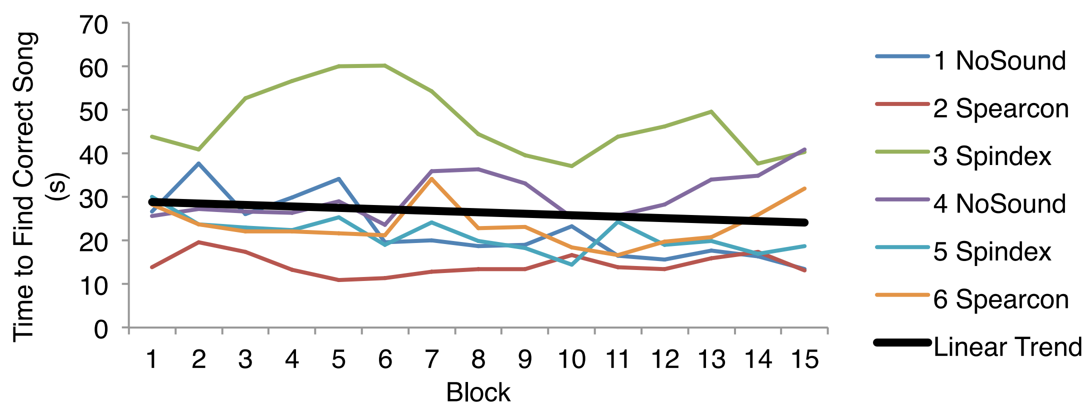


Figure 2.5 - Graph of the mean time to find a correct song for each participant across the 15 blocks. The thicker black line shows a linear trend line for the 6 participants.

List Search Summary. The results seen in the list searching data suggest that participants had increased their performance for the list search task and may have been able to continue increasing this performance. However, the decrease in standard deviations also suggests that participants may have been getting close to a plateau and their rate of learning may have slowed down.

2.3.2 Ball Drop Performance

For ball drop performance the data point of accuracy and number of balls released per block were collected. The descriptive data for these measures can be seen in Table 2 in Appendix I.

Accuracy. As seen in Figure 2.6 below the ball drop accuracy for the participants trended around 60 and 65 throughout the pilot. The calibration of the ball drop task at the beginning of each session may have limited the learning effects from being seen here as it controlled for learning effects of the single task on its own. The scalloped learning trend can be seen in the descriptive tables as well, with participants seemingly decreasing performance at the beginning of each session. A plateau could be argued for with this data but with the controlled calibration process this would conflict the results.

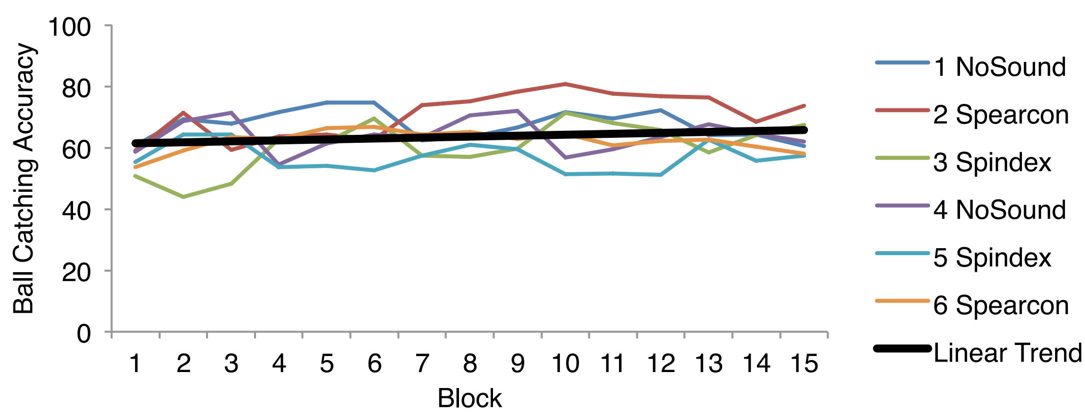


Figure 2.6 - Graph of the ball catching accuracy for each participant across the 15

blocks. The thicker black line shows a linear trend line for the 6 participants.

Balls Released. For the number of balls released the data and Figure 2.7 clearly show the increasing performance of the participants. Unlike the accuracy of the task, which was controlled via the calibration process, the number of balls released shows a clear learning effect over the period of the study. The standard deviation of the data seems to be decreasing towards the latter trials as well, suggesting a decrease in rate of change.

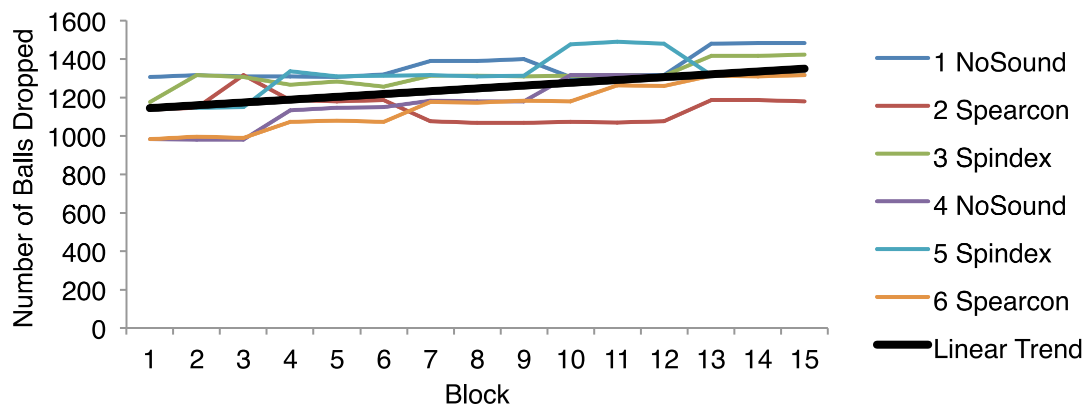


Figure 2.7 - Graph of the number of balls released each participant across the 15 blocks. The thicker black line shows a linear trend line for the 6 participants.

Ball Drop Performance Summary. The ball drop data suggested that learning was occurring over time for the participants, and that this change may be starting to slow down towards the end of the blocks. This data was a bit limited due to controlling the learning rate via the calibration process but still gave some information.

2.3.3 Visual Behaviors

For the fifteen blocks the descriptive data for time off rate, glance count rate, and average dwell length is seen in Table I.3. As is seen in the data one of the participants had a lot more variance than the other participants, however this was left in to ensure the full range of potential participants was considered.

Time Off Rate. The visualization of time off rate data can be seen below in Figure 2.8. As can be seen there and in the descriptive data, participants began to plateau soon after Block 10. While the rate of eyes off decreased slightly after it was minimal compared to the average decrease that occurred in the first 100 minutes of training. It can also be seen that some participants had much lower levels of time off rate from the start and had little variation over time, suggesting potential differences across participants.

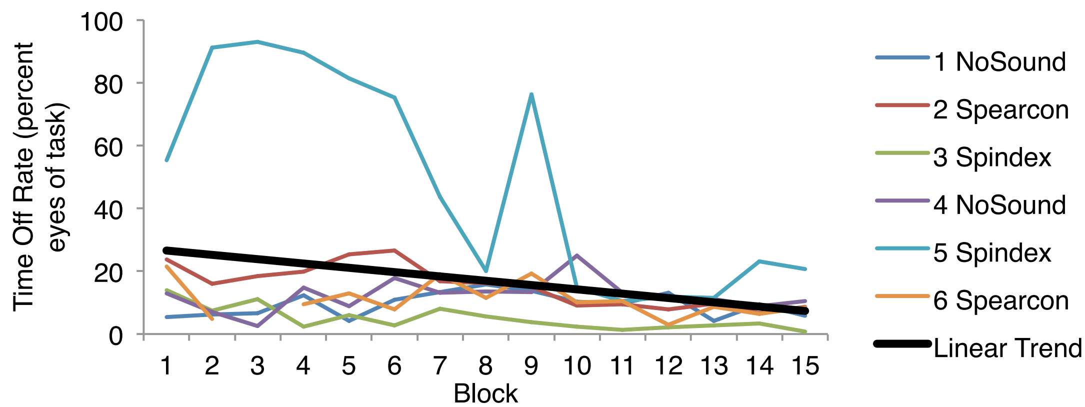


Figure 2.8 - Graph of the time off rate for each participant across the 15 blocks. The thicker black line shows a linear trend line for the 6 participants.

Glance Count Rate. The slope of glance count rate (mean number of glances per minute) as seen in Figure 2.9 tells a similar story to that of the time off rate. Participants seemed to be approaching a plateau towards the end of the study, with converging averages as they continued with practice. This would not be unexpected when considered with the time off rate data.

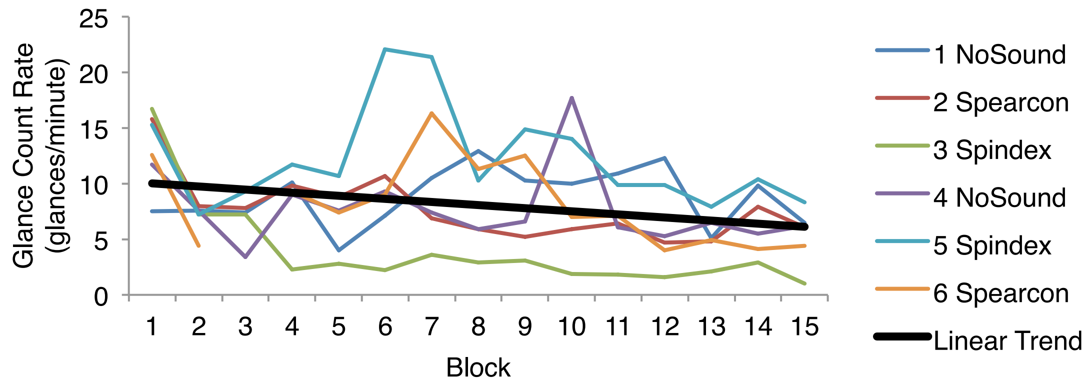


Figure 2.9 - Graph of the mean glance count rate for each participant across the 15 blocks. The thicker black line shows a linear trend line for the 6 participants.

Average Dwell Length. Figure 2.10 displays the graph of average dwell length over the 15 blocks. This data and the descriptive data in Appendix I show a decrease in rate of change and in standard deviation over time after about 100 hours of training.

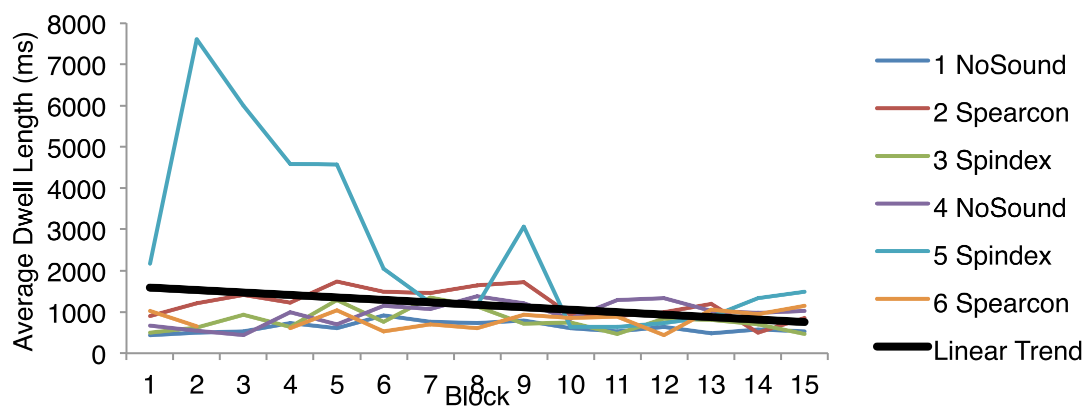


Figure 2.10 - Graph of the average dwell length for each participant across the 15 blocks. The thicker black line shows a linear trend line for the 6 participants.

Visual Behaviors Summary. The data for visual behaviors suggests that participants began to plateau around 100 minutes of training and having decreasing gains for practice beyond that point. While one participant contributed highly to this rate of change due to an initially high amount of visual behavior off the primary task it was assumed that these participants were a representative sample of those that would take part in the full study and therefore show what the full data might reveal.

2.3.4 Objective Workload

The descriptive data for the objective measures of mean HR and mean HRV for each block can be seen in Table 4 in Appendix I. The average heart rate across blocks can be seen in Figure 2.11. The trend suggests a slowly decreasing rate of change over time, with participants having a slightly lower average heart rate as the blocks go on. Similar to the heart rate data, the heart rate variability data also had a slow trend downwards, as can be seen in Figure 2.12. These slowly decreasing values suggest decreased objective workload over time, however the use of absolute values here over multiple days may not have shown accurate changes over the 5 sessions.

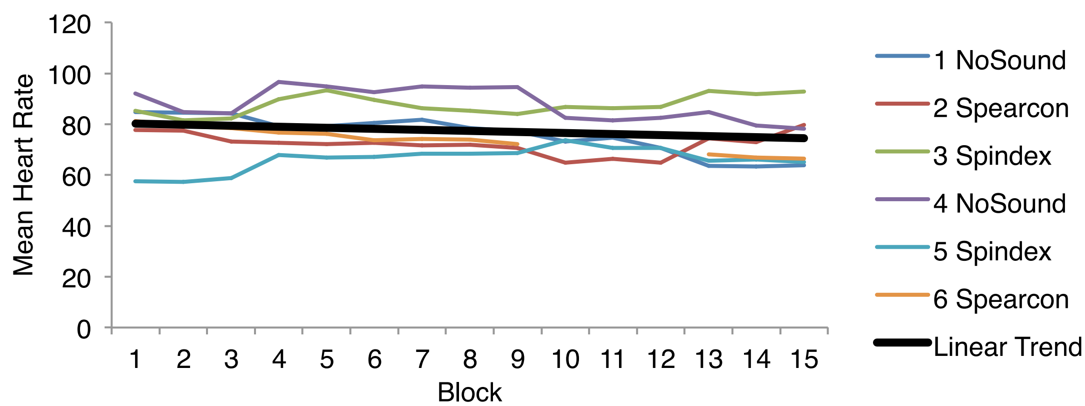


Figure 2.11 - Graph of the mean heart rate for each participant across the 15 blocks. The thicker black line shows a linear trend line for the 6 participants.

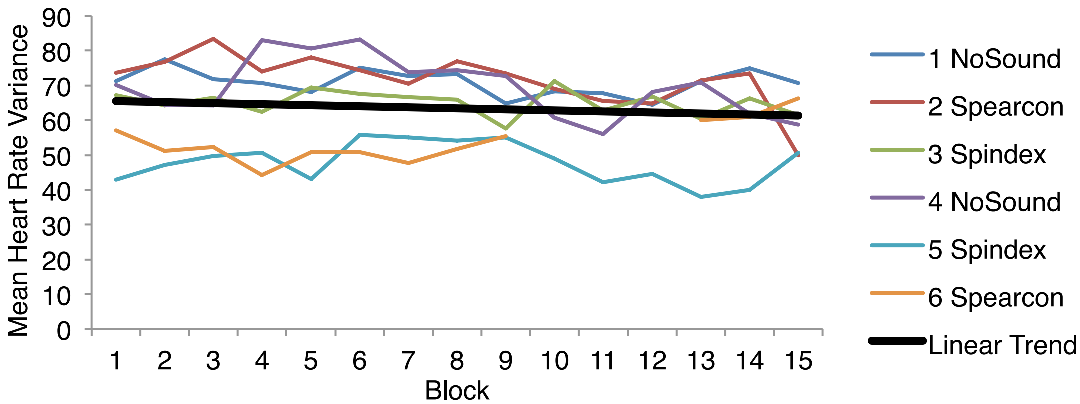


Figure 2.12 - Graph of the mean heart rate variance for each participant across the 15 blocks. The thicker black line shows a linear trend line for the 6 participants.

2.3.5 Subjective Workload

The descriptive data for each subscale and the composite workload across the 15 blocks can be seen in Table 5 in Appendix I.

Mental Workload. Figure 2.13 displays the graph for mental workload over the 15 blocks. As can be seen there and in the descriptive data a slow trend downwards was present for mental workload. Looking at the data though it seems that some participants found it to be very demanding throughout the study.

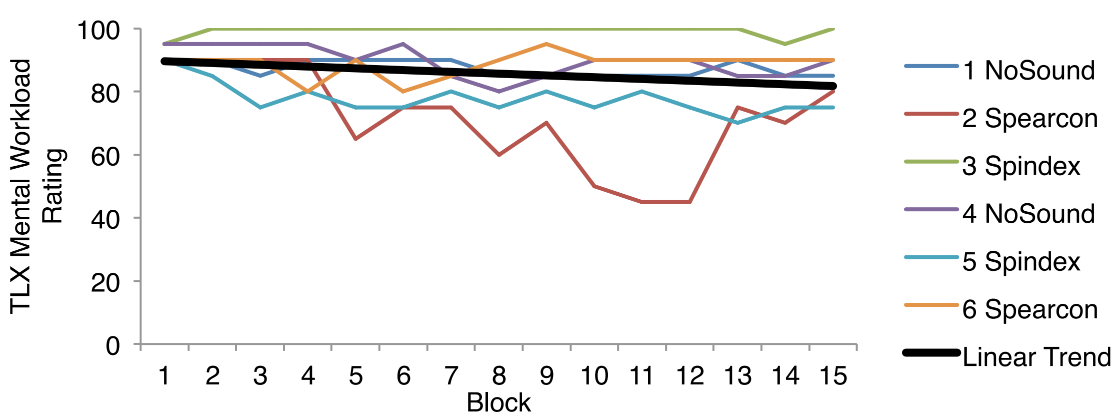


Figure 2.13 - Graph of the subjective mental workload rating for each participant

across the 15 blocks. The thicker black line shows a linear trend line for the 6 participants.

Physical Workload. Physical workload (Figure 2.14) showed the largest split between subjects than any other measure collected. The trend line shows a decreasing amount of workload but, as with mental workload some participants rated the workload high throughout the whole study and others had very low ratings throughout.

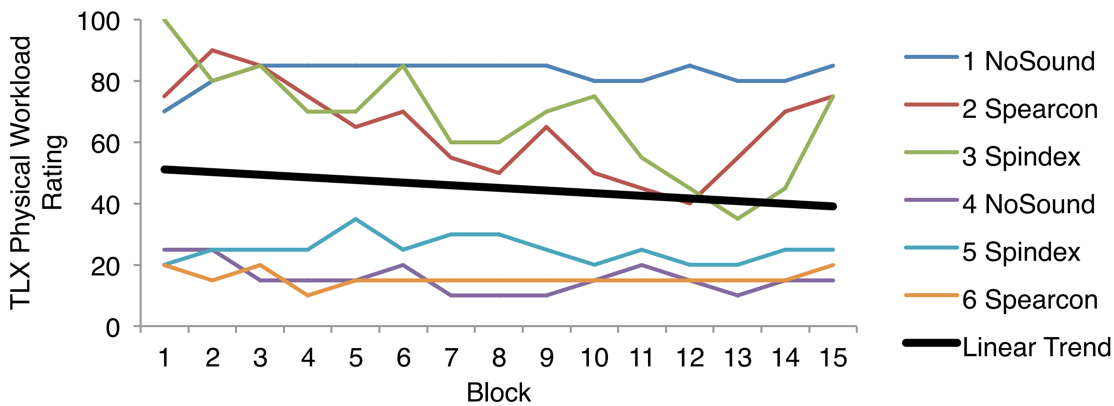


Figure 2.14 - Graph of the subjective physical workload rating for each participant across the 15 blocks. The thicker black line shows a linear trend line for the 6 participants.

Subjective Performance. Subjective performance as seen in Figure 2.15 was mostly consistent across the 15 blocks, with a slight trend downwards. This may have been due to a few participants mistaking the rating of low versus high for this measure part-way through the study due to misreading the question during the first few blocks.

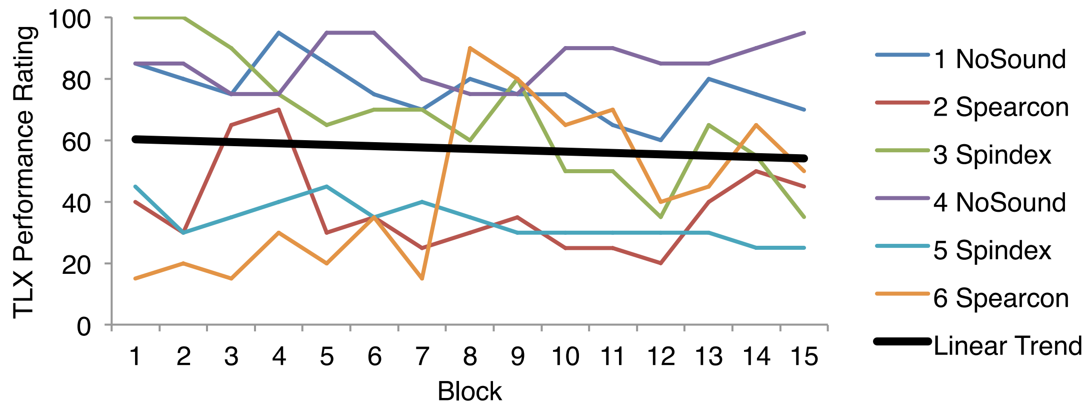


Figure 2.15 - Graph of the subjective performance rating for each participant across the 15 blocks. The thicker black line shows a linear trend line for the 6 participants.

Effort. Effort showed a clear downward trend as seen in Figure 2.16, even though a number of the participants still rated it fairly high throughout the pilot. The general decrease suggests less effort being needed to complete the tasks as the pilot went on. No plateau seemed to have been reached during this period although the ratings seemed to stay stagnant after about Block 8 looking at the descriptive data, suggesting 80 minutes was a point at which the decrease in effort may have slowed down.

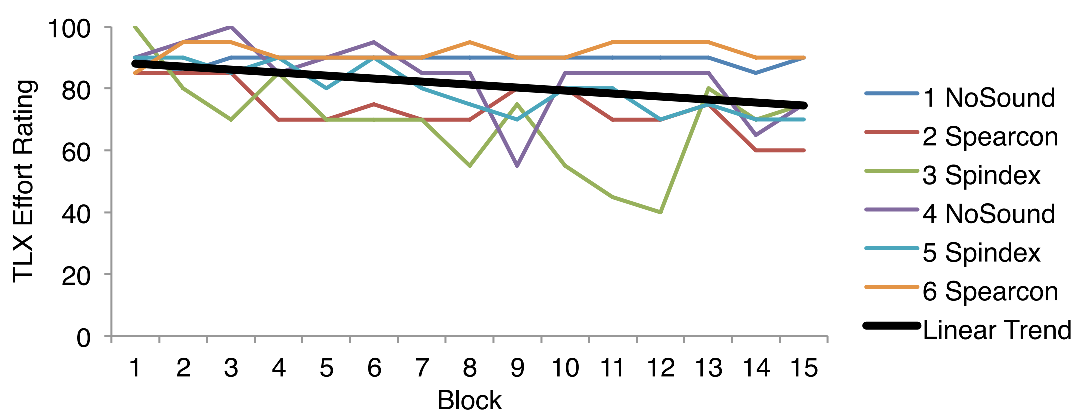


Figure 2.16 - Graph of the subjective effort rating for each participant across the 15

blocks. The thicker black line shows a linear trend line for the 6 participants.

Temporal Workload. For temporal workload there was little to no change over the 15 sessions, as seen in Figure 2.17. The descriptive data show a slightly decreasing standard deviation, suggesting the convergence of scores as time went on, but the lack of much change over time suggests it may not need to be variable in the study due to the changing of difficulty of the primary task over time to keep up with learning effects.

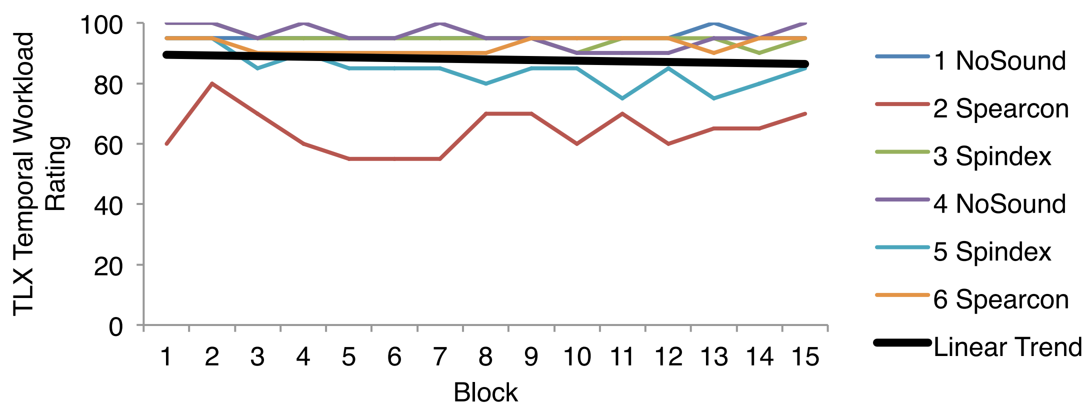


Figure 2.17 - Graph of the subjective temporal workload rating for each participant across the 15 blocks. The thicker black line shows a linear trend line for the 6 participants.

Frustration. As seen in Figure 2.18 frustration decreased a large amount over time in the pilot. Participants had fairly high ratings of frustration at the beginning, but the data shows a clear decrease in the trend line over time.

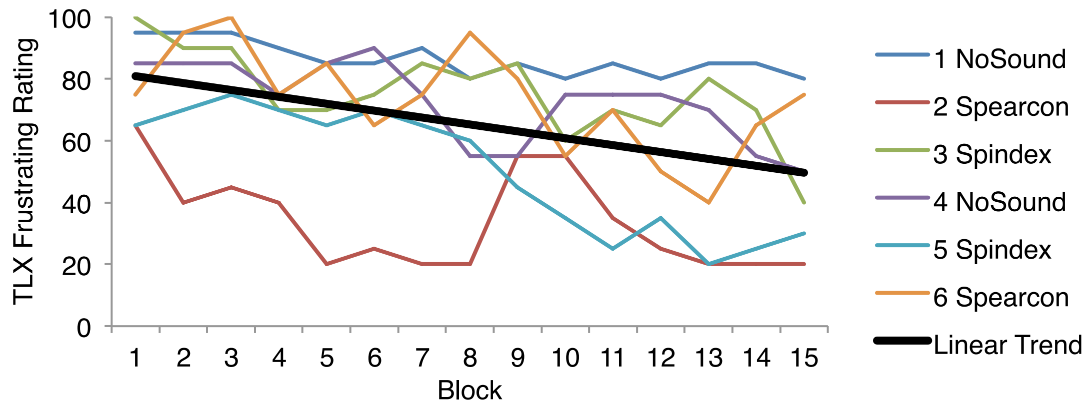


Figure 2.18 - Graph of the subjective frustration rating for each participant across the 15 blocks. The thicker black line shows a linear trend line for the 6 participants.

Composite Score. As seen in Figure 2.19, composite workload also had a slow decrease over time. However, looking at the averages per block in the descriptive data it seems that no real changes were seen after Block 5. This suggests that participants had a limited change overall in subjective workload after about 50 minutes.

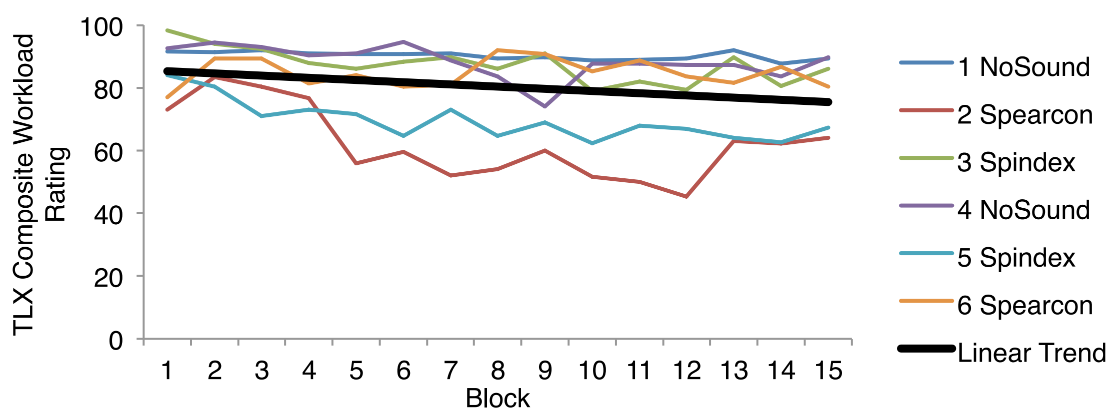


Figure 2.19 - Graph of the subjective composite workload rating for each participant across the 15 blocks. The thicker black line shows a linear trend line for the 6 participants.

Subjective Workload Summary. Overall the subjective workload data showed mixed results across the subscales. Some of the subscales such as mental, performance, and temporal showed no real changes during the whole 15 blocks, which could be due to changes in the difficulty of the primary task or issues in using the scale correctly. Others such as physical, effort, and composite showed a slow trending downwards in workload but the descriptive data suggested that no changes really happened in the scores in the latter half of the study. This suggests that about 80 minutes or so of training may be enough time as seen through subjective workload in this task.

2.3.6 Preferences

Table 6 in Appendix I displays the descriptive data for the measures collected in the survey on the 6-point, Likert-like scales across the 5 training sessions.

Effective At Search Task. Participants rated themselves as having higher effectiveness as the sessions went on except for a dip in ratings on the last session as seen in the descriptive data. This suggests Session 4 may be when they felt most effective and is displayed in Figure 2.20.

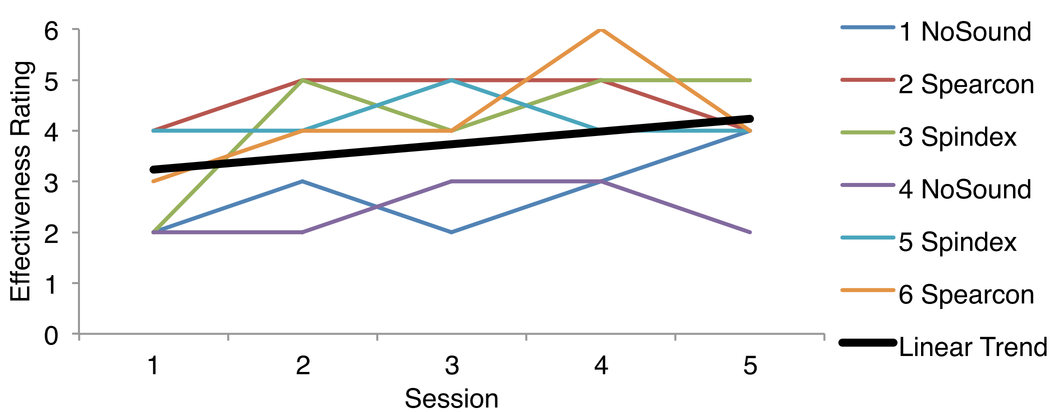


Figure 2.20 - Graph of the perceived self-effectiveness at the search task for each participant across the 5 sessions. The thicker black line shows a linear trend line for

the 6 participants.

Display is Effective. For the display being effective there seemed to be a change in the last session or two where participants felt it became a little more helpful at that point. This trend can be seen in Figure 2.21 below but the change only increased the average rating by about 0.5 as seen in the descriptive data.

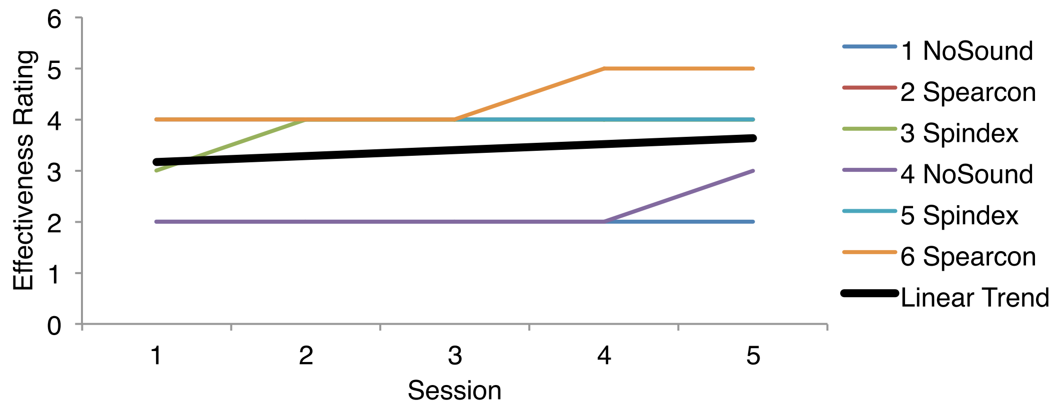


Figure 2.21 - Graph of the perceived effectiveness of the display for each participant across the 5 sessions. The thicker black line shows a linear trend line for the 6 participants.

Display is Functionally Helpful. For the display being helpful there was no real change over the five sessions as seen in Figure 2.22. Participants seemed to keep their opinion of the display throughout the study aside from a few changes. This suggests no changes may happen within this timeframe in regards to the measure.

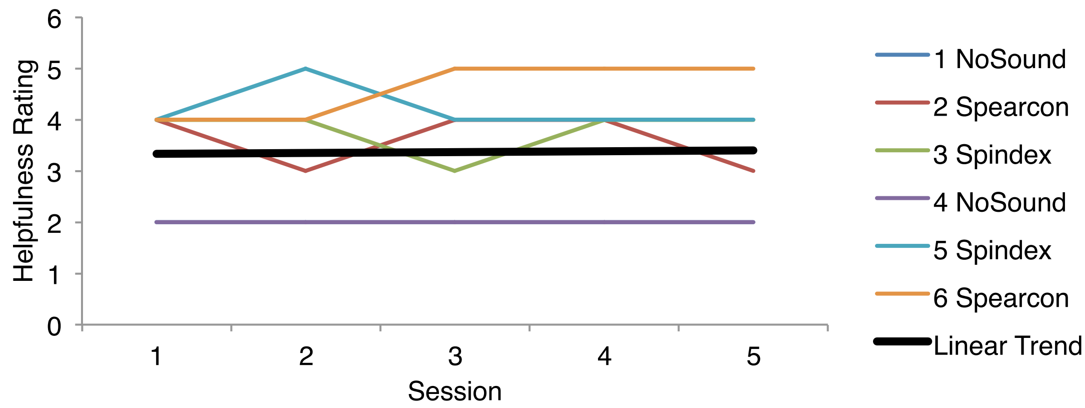


Figure 2.22 - Graph of the perceived helpfulness of the display for each participant across the 5 sessions. The thicker black line shows a linear trend line for the 6 participants.

Display is Annoying. As seen in Figure 2.23, participants' ratings of display annoyance seemed to decrease as time went on, with scores stabilizing between Session 4 and 5. This suggests that around 120 or less minutes of training is enough to decrease the initial amount of annoyance participants feel towards the display.

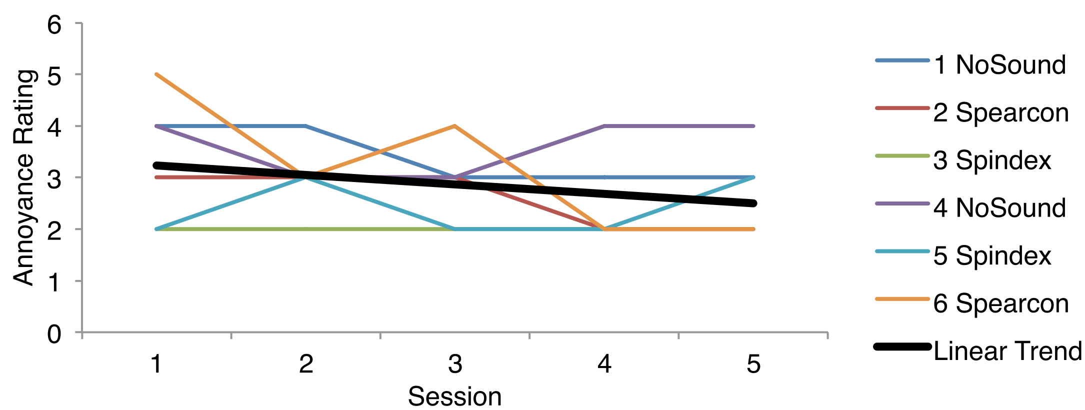


Figure 2.23 - Graph of the perceived annoyance of the display for each participant across the 5 sessions. The thicker black line shows a linear trend line for the 6 participants.

Effective At Primary Task. Viewing Figure 2.24 it seems that participants' rating of self-effectiveness at the search task increased slightly over time. Viewing the descriptive data it seems to have increased on average over the 5 sessions but only slightly.

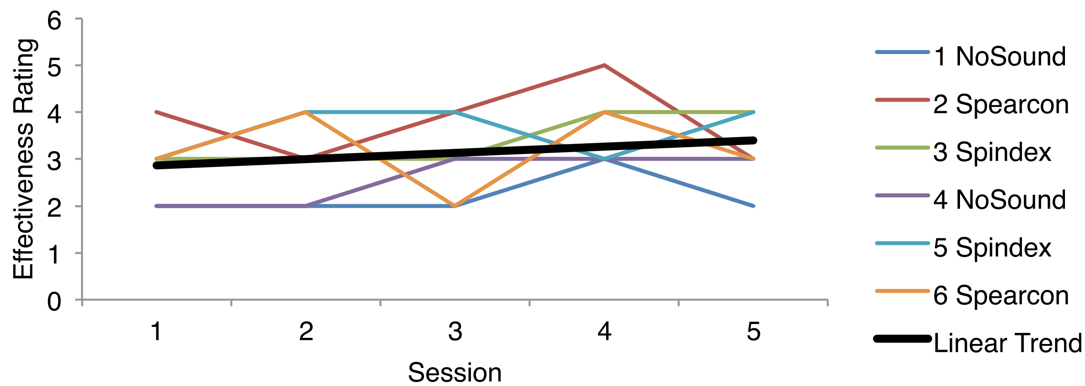


Figure 2.24 - Graph of the perceived self-effectiveness at the ball drop task for each participant across the 5 sessions. The thicker black line shows a linear trend line for the 6 participants.

Preferences and Perceived Performance Summary. Overall the participants' subjective ratings suggest a perceived improvement across the different measures. It seems that the ratings that changed over time began to change less across the last two sessions suggesting that 10 minutes or maybe less may be enough training to see similar effects of perceived change.

2.4 Pilot Discussion

RQ1.1. In the pilot, Research Question 1.1 was asking if practice would affect participant performance on the secondary task, visual behaviors, workload, and perceived performance/comfort. Hypotheses 1.1.1 - 1.1.4 focused on whether practice affects each dependent measure and if so, which way the data pointed.

Regarding H1.1.1, which predicted increases in performance on the search task, the data pointed towards learning occurring. The general trend across the 15 blocks was an increase in the number of searches and percent accuracy, with a decrease in average search time. This suggests participants were improving at the search task as time went on, supporting the hypothesis.

For H1.1.2 the data seemed to support the hypothesis that workload would decrease as training went on. In regards to subjective workload, most of the measures decreased across the 15 blocks, with only performance (which was due to some confusion) and temporal workload (potentially due to the increasingly more difficult ball drop game caused by the calibration process) seemingly not changing across the pilot. For objective measures heart rate seemed to decrease slightly, while heart rate variance mostly stayed the same. This suggests, that even with the increasingly more difficult ball drop game due to calibration, that workload was decreasing overall, suggesting it was taking participants less demand to complete the dual task over time.

H1.1.3 was focused on addressing the issue of no workload change during the study if it came up, particularly if it was due to the increasingly difficult primary task. While this was not seen in the NASA-TLX data, the participants' opinions of the display and performance on the different tasks also pointed towards learning effects. This was seen through higher effectiveness ratings of themselves at both the primary and secondary tasks as well as better scores for effectiveness of the display and annoyance of the display at he end of the study than at the beginning. This again points to decreasing workload and increasing comfort with the displays over time.

For H1.1.4 the results from the visual behaviors also were seen to support the idea that learning was occurring, with the data showing trends of less looking towards the phone as practice with the displays increased. This was seen through decreased number of glances, number of long glances, and percent time eyes off the primary task as time went on. This suggests participants were getting used to the interactions and needed to look less often at the phone to get the same amount of data, therefore learning how to complete the dual task more efficiently.

Across RQ1.1 the results of the pilot suggested that learning effects were occurring across the measures. This supports the hypotheses made and confirms that participants are able to improve their performance on these tasks over time. This suggests that studying these tasks with more participants could reveal some differences between conditions.

RQ1.2. Research Question 1.2 was aimed at determining whether or not participants would reach sufficient expertise in the pilot period and how long that would take if they did. H1.2.1 stated that changes in task performance, workload, perceived performance/comfort, and visual behaviors would begin to slow down as participants had more practice. The trends of the pilot suggest that this may be occurring for some of the measures. In particular, subjective measures of effort and composite NASA-TLX workload seemed to change more slowly after about 80 and 50 minutes respectively. Visual behaviors also seemed to slow in change after about 100 minutes of training across the three measures. The measures of ball drop changed throughout as would be expected due to the constant changing of difficulty with calibration, and while participants were still improving in regards to the list search task the end of that learning process is unknown.

Timing Decision. The results from the pilot pointed at RQ1.1 being answered in a way pointing towards participants having learning effects in the time provided, and somewhat for RQ1.2 suggesting that there may be a level to which participants can be trained within the period given. This answer for RQ1.1 suggests that the full study could be beneficial and gain new insights for the area of multimodal displays, meaning the study should move ahead. For RQ1.2, the data pointed to about 100 minutes of training being the point at which changes in some measures seemed to begin to slow. For this reason 100 minutes of training was chosen as the minimum time for the full study.

In addition to the planned data gathered from participants during the pilot a few other pieces of data were also determined and used to help shape the full study. First, the 10 minute time period for each session was found to be slightly too long. Participants were complaining of fatigue within each block and multiple participants stated that slightly shorter times would be easier to deal with. To address this the training blocks were shortened in the full study. In addition, the lack of any differences seen trend-wise for heart rate and heart rate variance were thought to possibly be due to within-subject variation of heart rate variables over the sessions. To address this the full study also was planned to have baseline heart rate variable measurements at the beginning of each session to be used to create a percent difference score calculation for each session's data, which would be easier to compare across sessions for within-participant data as well as between-participant data. Finally, the participants also found the ball drop task to be extremely difficult during the pilot. Multiple participants felt the difficulty was limiting their ability to perform the dual tasking in an effective manner and therefore the percent accuracy of the task was released by ten percent for the calibration process.

CHAPTER 3: FULL STUDY

The trends in the pilot pointed towards differences in performance and behaviors over time due to training on the list search task. Based on these trends and the time it took for the participants to reach a basic level of training it was determined that in the experiment participants should get at least 100 minutes of practice in the dual task training blocks to be able to see similar training effects. To make this fit within a reasonable time for participants the training was broken into 13 8-minute blocks (totaling 104 minutes). The shortening of blocks was also done in response to pilot feedback of blocks being too long. To test the effects of this training within an applied setting, the study also included the testing blocks where participants performed the dual task in a driving simulator at the beginning and end of the study to see the difference in abilities in a real world task.

3.1 Research Questions and Hypotheses

RQ2.1. The first research question in the experiment, RQ2.1, was similar to RQ1.1 and served the purpose of a manipulation check: **Did the amount of practice given affect participant performance in the dual task situation as seen in the pilot?** This was determined by looking at how the visual behaviors, workload, and the dual-task performance (accuracy and speed) across the training blocks and sessions changed and comparing these trends to that of the pilot. This was expected to be particularly evident in the driving blocks due to calibration changes that occurred in the training blocks, therefore continually increasing difficulty of the dual task. To investigate this question the following hypotheses (H) were forward to test.

H2.1.1 – Participants' secondary task performance would increase as they got more practice with the displays. This would be seen through increased accuracy and decreased search times for the secondary task and was expected due to training effects, with practice increasing performance on tasks over time which was seen in the pilot.

H2.1.2 –Primary task performance would also increase as participants gained practice. This would be seen through better driving performance in the driving blocks or better ball catching accuracy or number of balls released for the Ball Drop task. As with the previous hypothesis, this would be expected to occur if a practice effect takes place.

H2.1.3 – Participants would decrease their visual behaviors off the primary task as practice with the displays increases. This would be seen through lower levels of percent time eyes off the primary task, number of glances off task, and average dwell time off the primary task. This should occur as practice will get the participants accustomed to the display they are using and help them to form a more effective way to approach their task sharing regarding visual behaviors.

H2.1.4 – Workload as seen through measures of percent change of mean HR and HRV and NASA-TLX scores would decrease as participants got more practice with the display over the sessions of training. This was hypothesized, as it should take participants less workload to complete the tasks as they get more practice.

H2.1.5 – Participants would report feeling more comfortable and have higher subjective performance scores with their assigned displays as they go through more training sessions. Again, this was expected due to the increased time with the displays and therefore expected increase in comfort.

RQ2.2. Research Question 2.2 was aimed at investigating the differences in the effects of training between the conditions: **How does practice differently affect performance of participants in the three conditions?** This was similar to R2.1 but was aimed at looking for the differences between the three conditions of displays. This was determined by looking at the visual behaviors, workload, and the dual-task performance (accuracy and speed) throughout the training blocks across the conditions. Again this was expected to be more prevalent in the driving blocks. To investigate this research question the following hypotheses were put forward to test.

H2.2.1 – Practice would unequally increase the secondary task performance of those using the auditory cues as seen through accuracy and time. This was expected as the theories of multitasking and multimodal interaction state that multimodal displays should lead to better time sharing/task performance (Wickens, 2002) and because it was expected that those using the visuals-only condition would benefit less from practice due to their expected initial higher levels of ability.

H2.2.2 – The performance on the primary task of ball drop was hypothesized to also unequally increase for those using the auditory cues as seen through balls released and accuracy. Although the number of balls released was determined in the calibration process, differences in performance between conditions could also be a way to point to differences in learning rates.

H2.2.3 – Those in the auditory conditions would have larger decreases in their visual behaviors over time of average percent eyes off task time, numbers of glances, and average glance duration off the primary task than those in the visuals only condition. According to the theories that drive the use of auditory cues participants should be able to continue the search task while using their visual attention to attend to the primary visual-manual task instead of switching visually between the primary and secondary tasks, as the visuals-only condition requires. The visuals-only condition participants have a limit to which they can decrease their use of visuals.

H2.2.4 – Participants' subjective and objective workload on the multitasking situation would decrease throughout the training phase more so for those with the auditory cues than those in the visuals-only condition. This was hypothesized because according to MRT the multimodal cues should allow for an eventual lower level of workload than using only visual cues, therefore decreasing both objective and subjective workload as seen through lower NASA-TLX workload ratings and lower mean heart rate.

H2.2.5 – Participants in the auditory conditions would rate their perceived performance and comfort on the dual task situation, in particular the search task, as increasingly better throughout the training phase. Experience with the task should increase participants' perceived performance as seen in other work, displayed through increased perceived performance and comfort scores.

RQ2.3. The final research question for the study was regarding dependent measures during the testing blocks across the conditions: How would performance compare across the display types in the testing condition? As with the other research questions this was determined by looking at the measures of visual behaviors, workload, and the dual-task performance (accuracy and speed) but only during the testing block. To investigate this question the following hypotheses (H) were put forward to test.

H2.3.1 – Secondary task performance as measured through accuracy and time was hypothesized to be higher in the auditory cue conditions than in the visuals-only condition. This was expected due to similar reasons already discussed of the multimodal displays allowing for better time-sharing and therefore increased performance on the search task than compared to a single modality of use for both tasks.

H2.3.2 – As with the secondary task performance, driving performance was also hypothesized to be better for those in the auditory conditions due to better task sharing. This was expected to be seen as measured through improved lane keeping, and following distance measures.

H2.3.3 – Those in the auditory conditions would have lower average percent eyes off task time, glance rate, and average glance duration off the primary task than those in the visuals only condition. Similar to the other visual behavior hypotheses this was expected as MRT and previous work suggests that participants with the auditory cues should be able to continue the search task using the auditory information without taking their eyes off of the primary visual-manual task as much as the visuals-only condition requires, particularly once they were used to using such displays.

H2.3.4 – Subjective and objective workload on the multitasking situation would be lower for the two auditory conditions than the visuals-only condition. This was because MRT suggests that the multimodal cues should allow for a lower level of workload than using only visual cues, therefore decreasing both objective and subjective workload as seen through lower NASA-TLX workload ratings and lower levels of workload seen through HR measures.

3.2 Methods

3.2.1 Participants

A total of 55 participants took part in the study, however three were thrown out due to not completing all four sessions and only the remaining 52 that completed all sessions are discussed hereafter. The participants were recruited via SONA or through word of mouth and were required to have a valid driver's license in the United States for a minimum of two years; report normal or corrected to normal vision, hearing, and mobility; and avoid performing any strenuous exercise or caffeine intake for two hours prior to participating in the study. Participants recruited through SONA received study credits for their time (1 credit for each hour).

The 52 participants who completed the study (31 males, 21 females) had an average age of 22.23 (S*D* = 4.85), and had been driving for an average of 5.78 years (*SD* = 4.49). Participants reported driving an average of 5 hours per week (*SD* = 3.95), with 39.37% of that being on the highway (*SD* = 27.75). Participants also reported using technology for 28.92% of that driving time (*SD* = 31.02).

3.2.2 Apparatus and Materials

List Search Displays. The same three conditions with the same types of displays as in the pilot were present in the full study, including the visuals-only condition (No-Sound), visual + Spearcon + TTS (Spearcon) condition, and visual + Spindex + TTS (Spindex) condition. Participants only used one of these displays throughout the entire study, as condition was randomly assigned to them.

List Search Task. The participants used the same phone and search task as in the pilot. Participants still chose the hand they wanted for the secondary and primary tasks, using only one hand for the driving task. Each search period was 8 minutes long so as to adjust according to the updated training block timing.

Performance on the search task was again tracked through accuracy, number of songs searched in each block, and the time it took participants to find a correct song. These data were calculated and stored by the phone for the training and driving blocks independently and then uploaded following data collection to a database.

Driving Simulation. To determine the potential impact on driving performance participants drove in two baselines (single tasks only) and three testing blocks (dual tasks) throughout the study. The baselines were collected as a means to control for individual differences between participants. In Session 1 each participant drove a 2-minute baseline drive on a straight, 2-lane highway followed by an 8-minute dual task drive on the same type of road. Then in Session 4 participants drove another 2-minute baseline and 8 minute straight drive as well as an additional 8-minute curvy road on a 1-lane (in each direction) highway. All of the drives were all a following task where participants were instructed to follow a lead car. The lead car intermittently changed speeds with the average speed being 50 MPH and a standard deviation of 10 MPH. The participants were instructed to follow the lead car at a distance of 50 feet, which was displayed to them at the start of the drive. In all of the drives there was traffic going in both directions but never interacting with the lead vehicle or the drivers' car.

These driving scenarios were performed on a National Advanced Driving Simulator MiniSim. The simulator was composed of three 42" plasma displays with 1280x800 pixel resolutions and a 130-degree field of view to display the graphics of the simulator, and a 2.1 sound system to play environmental sound effects. The simulator was controlled using a real steering wheel and brake and gas pedals. A smaller LCD built into the simulator was used to show the speedometer and tachometer. The scenarios were created using the NADS internal ISAT creation tool. The setup for the driving blocks can be seen in Figure 3.

Figure 3.1 - Participant performing the driving task while searching. Also pictured is the eye-tracking and heart rate hardware and software.

Measures of driving performance collected in the study were measures of longitudinal control, namely mean and mean standard deviation of distance gap (front gap) – defined as per SAE J2944 as the distance (in feet) between the driver's front bumper and the lead car's rear bumper – and mean and mean standard deviation of lane position (SDLP) – defined as per J2944 as the distribution of the lateral lane position (in feet) (SAE, 2013). During each drive, the simulator collected and stored these driving measures at a rate of 60 Hz. As is discussed later in the document, values from these measures were calculated based on percent change from the baseline drives to control for any differences between participants based on their driving abilities prior to the study.

Driving Block Visual Behaviors. The visual behaviors of the participants during the driving blocks were measured via a three camera Smart Eye Pro tracking system. The SmartEye system was set to gather data at 60 Hz using the dark pupil method. As with the training blocks the eye tracker software was used to create a 3D model of the primary task screens to determine when a participant was looking at the driving task or not. To follow the same process as in the training blocks, any frames when the participant was not being tracked as being on the primary task, was reported as time with eyes off the road.

From this, the same data values as in the training blocks were calculated including: percent dwell time off-task (calculated by dividing the number of frames where the participant was not on the primary driving task by the number of frames collected for that drive); mean glance frequency off-task per minute (average number of glances off the primary task per minute); and mean glance duration off the primary task (the average time in ms that participants spent off the primary driving task during a glance). However, while these measures would work, eye-tracking data was also collected during the baseline drives. As is described further in the document this data was used to create mathematical difference scores to control for any loss of eye-tracking by subtracting the measure for the baseline drive eye-tracking data from the driving block eye-tracking data.

Visual-Manual Training Task. For the training blocks participants performed the same ball drop game as in the pilot. The visual manual vigilance task consisted of the same seven vertical columns from which the balls fall, and a paddle at the bottom to catch the balls. The participants controlled the paddle via the arrow keys on the same hardware. As before, the goal was to catch as many balls as possible. The dropping rate of the game was calibrated individually for each participant at the beginning of each session so that each participant caught between 85-95 percent of the balls that were released, and balls could start a maximum of two columns away from where the previous ball was released. Performance on the ball drop task was tracked by the software via accuracy (determined by the number of balls caught divided by the total number of balls released during each block), number of balls released during the block. This was data was stored in a log file after each session.

Training Block Visual Behaviors. The visual behaviors of the participants were measured using the same set of FaceLAB 4 fixed eye trackers as in the pilot. The system's two cameras were set to use the dark iris or dark pupil depending on best reading for the individual participant to gather data at 60 Hz during the study and the eye-tracking was used to determine when the participant was looking at the ball drop task and when they were looking down at the secondary search task. Due to loss of tracking at certain angles, particularly when participants look away from the primary task, any frames when the participant was not being tracked as being on the primary task were reported as off task.

As in the pilot, these data were stored in log files and synced with the other data being collected. Eye-tracking values as defined by SAE J-2396 (Lamble, et al., 1999) were calculated to look at the visual behaviors including: percent dwell time off-task (calculated by dividing the number of frames where the participant was not on the primary task by the number of frames); glance frequency off-task per minute (number of glances off the primary task per minute); and mean glance duration off the primary task (the average time in ms that participants had of the primary task).

Subjective Cognitive Load. Subjective cognitive load was measured through NASA-TLX in the same way it was in the pilot. The TLX measured the six subscales of workload including effort, temporal demand, physical demand, frustration, performance, and mental demand. Data from the TLX survey was saved on a computer once completed. As in the pilot both the single numerical value of total workload and each raw subscale were pulled to investigate each of the individual factors.

Objective Cognitive Load. As in the pilot the physiological workload measures of HR and HRV were collected. The data was collected with the same NeXus-10 physiological monitoring and biofeedback platform as in the pilot, gathering EKG data at 256 Hz from the participants via a modified lead II configuration. As in the pilot, the physiological system collected and stored the data, which was matched via timestamp along with the primary task. However, unlike the pilot, percentage-change difference scores were calculated to allow for fair comparison across sessions and participants, which is hard to do with the raw data. To do this, participant's baseline HR measures for each session were subtracted from the data for each block for the same measure, divided by the baseline measurement, and multiplied by 100 to create a percentage-change score. So for Session 1, Block 1, mean HR, the equation would be ((Block 1 - Baseline 1)/Baseline 1)*100. This was done for both measures across all 15 blocks and resulted in 15 difference score values for each measure.

Other Measures. Participants completed the same two types of surveys over the duration of the study: a demographics questionnaire (Appendix B) given during the first session of the study; and a preferences questionnaire (Appendix C) given for each session regarding perceived performance on the cell phone and visual manual tasks for that session.

3.2.3 Procedure

The procedure was similar to the pilot aside from the number of sessions that the participants took part in, the time length of the blocks, and the addition of the driving blocks. Based on the findings from the pilot it was determined that at least 100 minutes of training was needed. Due to reports of fatigue from participants in the pilot, the block times were decreased to eight minutes in length. In total the study included thirteen 8-minute training blocks over four sessions, giving participants 104 minutes of training time. In addition to these training blocks there were also three 8-minute driving blocks and two 2-minute baseline drives. The timing and spread of these blocks and the setup and questionnaires are displayed in Table 3.1.

Table 3.1 - Session and block layout for the full experiment. Note that blocks are not to scale regarding visual size of block and time it took.

<table>
<tr><th>Session 1</th><th>Session 2</th><th>Session 3</th><th>Session 4</th></tr>
<tr><td rowspan="2">Study screening and consent</td><td>HR setup</td><td>HR setup</td><td>HR setup</td></tr>
<tr><td>Ball drop setup</td><td>Ball drop setup</td><td>Ball drop setup</td></tr>
<tr><td rowspan="3">Study and display intro</td><td>Ball drop Block 2</td><td>Ball drop Block 7</td><td>Ball drop Block 12</td></tr>
<tr><td>TLX</td><td>TLX</td><td>TLX</td></tr>
<tr><td>Ball drop Block 3</td><td>Ball drop 8</td><td>Ball drop Block 13</td></tr>
<tr><td>HR setup</td><td>TLX</td><td>TLX</td><td>TLX</td></tr>
<tr><td>Driving setup and baseline</td><td>Ball drop Block 4</td><td>Ball drop Block 9</td><td>Driving setup and baseline</td></tr>
<tr><td>Drive Block 1</td><td>TLX</td><td>TLX</td><td>Drive Block 3</td></tr>
</table>

TLX	Ball drop 5	Ball drop Block 10	TLX
Ball drop setup	TLX	TLX	Drive Block 3
Ball drop Block 1	Ball drop Block 6	Ball drop Block 11	TLX
TLX	TLX	TLX	Questionnaires
Questionnaires	Questionnaire	Questionnaire	Debrief

During the first session participants first confirmed they met the screening details and then read through and signed the consent form if they wanted to participate. Participants were then given the study description form to brief them on the study plan and details. Participants were then introduced to the phone interaction and the display they would be using in the study. Next, they were walked through the Georgia Tech Simulator Sickness Screening Protocol (GTSSSP) as described in Gable and Walker (2013). In the SSSP participants fill out a questionnaire regarding their current physical feeling, drive for two minutes in the simulator, and then fill out the same questionnaire again. If during the drive the participants report feeling sick or if after the drive the questionnaire reveals an increase in sickness feeling, they are to be released from further participation and given credit for participating. No participants were found to have simulator sickness. Participants were then fitted with the heart rate and eye-tracking system in the simulator, followed by an explanation of the driving task and a 1-minute heart rate baseline. They then completed the baseline-driving course alone, followed by the straight drive while doing the secondary task. Afterwards they competed the NASA-TLX for the dual task drive. The participants then moved over to the ball drop task where they were fitted with the eye trackers, completed the ball drop calibration, and then a single block of the ball drop task and the associated TLX survey. They were then given the demographics questionnaire and the session questionnaire for Session 1.

During Sessions 2 and 3 participants were first fitted with the heart rate system, had the heart rate baseline measured, and had the ball drop eye trackers calibrated. This was followed by a calibration session with the ball drop and then five sessions of ball drop and the associated TLX surveys. At the end of both sessions participants again did they session survey, with Session 3 including one additional question regarding their strategy in completing the dual task.

For Session 4 participants again were fitted with the heart rate system and did the 1-minute baseline measure, followed by the ball drop eye-tracking fitting and task calibration. After two blocks of ball drop and the associated TLX measures, participants once again moved over to the driving task. Here they completed another 2-minute baseline drive followed by two, 8-minute drives, the first being the straight drive and the second being the curvy road. After each of these drives participants completed the TLX survey. Finally the participants were given the sessions questionnaire and then the debrief form (Appendix H).

3.2.4 Data Organization, Design, and Analysis

The independent variables for the experiment were the amount of training participants received, and the three conditions in which participants were randomly placed in. Many of the dependent variables were the same as in the pilot, including: visual manual training task performance (accuracy, balls caught, and balls released); list search performance (accuracy, songs found, and time to find a song); visual behaviors for the training blocks (percent time eyes off the primary task, number of glances off task per minute, and average dwell time off the task); heart rate measures (percent change in mean HR and HRV); NASA-TLX scores; reported levels of annoyance; preferences; and perceived performance. In addition to these variables the measures of driving performance (percent change in mean and percent change in standard deviation of distance gap, and percent change in mean and percent change in standard deviation of SDLP) as well as the visual behavior measures in the simulator (change in percent time eyes off the primary task, change in number of glances off task per minute, and change in average dwell time off the task were recorded).

After collection and calculation, all of the measures except for the survey data were analyzed for outliers and had them removed (the survey Likert-like data was not analyzed for outliers due to limited range). The outliers were calculated within each condition within each block, which was done to ensure that no cross condition or cross block (learning effect) differences were used to inform the determination of outliers. Outliers were defined as any data value less than or greater than the outlier range, calculated by determining the interquartile range (IQR) multiplied by 1.5 and then subtracting this value from Q1 and adding the value to Q3.

Driving Block Analyses. For the driving blocks a series of 3x3 (conditions x driving block) mixed-model ANOVAs (with Huynh-Feldt corrections to address sphericity) were used to test for differences between the three driving blocks across the three conditions. These were followed up by t-test post-hocs when significant (paired for the main effect of drive; independent sample for the main effect of condition). All analyses were performed or corrected to alpha = 0.05, with LSD corrections applied to the post-hocs to correct for any type-2 error due to this being the most effective correction when three tests are being performed.

Training Block Analyses. As in the pilot study, the data in the full data set had a scalloped learning pattern and the data was therefore analyzed via training session instead of via training block to decrease noise in the analyses. This meant that the dependent measures from the training blocks were averaged across sessions and then analyzed through a series of 3x4 (condition x session) mixed ANOVAs (with Huynh-Feldt corrections to address sphericity). When significant differences were found these were followed up by t-test post-hocs (paired for the main effect of block; independent sample for the main effect on condition) to determine where the significant differences between the conditions, blocks, or any interactions laid. All analyses were performed or corrected to alpha = 0.05, and to correct for any type-two error corrections were applied to the post hoc tests. For post-hocs of the main effect of condition, which included three t-tests, LSD corrections were applied, and for the six t-tests used as post-hocs for the main effect of training sessions Bonferroni corrections were applied, lowering the alpha level to .008.

To investigate significant interactions, a set of seven one-way ANOVAs were performed across all combinations of Condition x Session and Session x Condition, with Bonferroni corrections decreasing alpha to .007 for those tests. If any of these one-ways were found to be significant they were followed up by t-tests for that specific combination and corrected. These corrections meant that in the case of any Session x Condition combinations there were three t-tests, which was corrected with LSD (alpha = .05), and for any combination of Condition x Session with six t-tests, it was corrected for using Bonferroni (alpha = .008).

Survey Data Analysis. Subjective data from the surveys was analyzed in a similar manner to the driving blocks via 3x4 (conditions x session) mixed-model ANOVAs (with Huynh-Feldt corrections to address sphericity). These were followed up with t-test post-hocs when significant (paired for the main effect of session; independent sample for the main effect on condition). All analyses were performed or corrected to alpha = 0.05, with LSD corrections applied to the condition post-hocs and Bonferroni to the session post-hocs to correct for any type-2 error.

3.3 Driving Block Results

3.3.1 Driving Blocks List Search Performance

For measures of list search task performance during the driving task the three measures of number of trials, percent accuracy, and average time to find a correct song were used. The descriptive data from these measures can be seen in Table J.1 in Appendix J.

Number of Searches. For the number of trials in the driving blocks the ANOVA revealed a significant main effect of driving block. This was investigated via paired t-tests and revealed that participants performed more total searches in Drive 1 than Drive 3, and in Drive 2 than Drive 3. No significant difference was found for the main effect of condition, nor was there a significant interaction. These analyses can be seen in Table 3.2 and the data are visualized in Figure 3.2 below.

Table 3.2 – Analysis table for the mean number of search trials for each condition across the three drives. Significant differences are marked with "*".

Number of Searches

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Drive	9.56	1.53, 72.09	.001*	.169

Main effect of Condition	0.89	2.00, 47.00	.419	.036
Interaction	1.51	3.07, 72.09	.218	.060
Post-hoc t-test Results	*t*	df	*p*	
Drive 1 – Drive 2	1.79	49.00	.079	
Drive 1 – Drive 3	3.59	49.00	.001*	
Drive 2 – Drive 3	3.73	49.00	.001*	

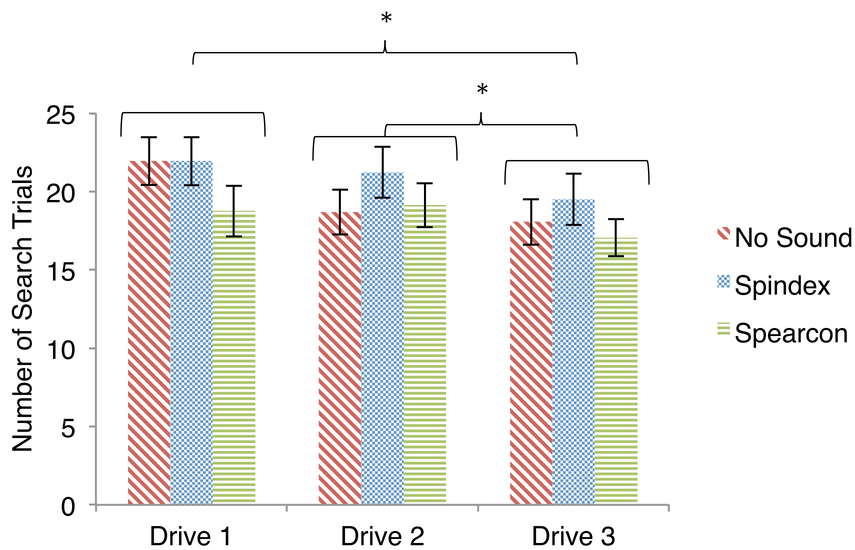


Figure 3.2 - Graph of the mean number of search trials for each condition across the three drives. Standard error is shown via error bars and significant differences are marked with "*".

Percent Correct. For the percent correct songs chosen in the driving blocks the ANOVA revealed there to be no significant main effect of driving block, no significant difference for the main effect of condition, nor a significant interaction. These analyses can be seen below in Table 3.3 and Figure 3.3.

Table 3.3 – Analysis table for the mean percent correct selections for each condition across the three drives. Significant differences are marked with "*".

Percent Correct

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Drive	2.03	1.99, 89.60	.137	.043
Main effect of Condition	0.69	2.00, 45.00	.507	.030
Interaction	0.84	3.98, 89.60	.137	.043

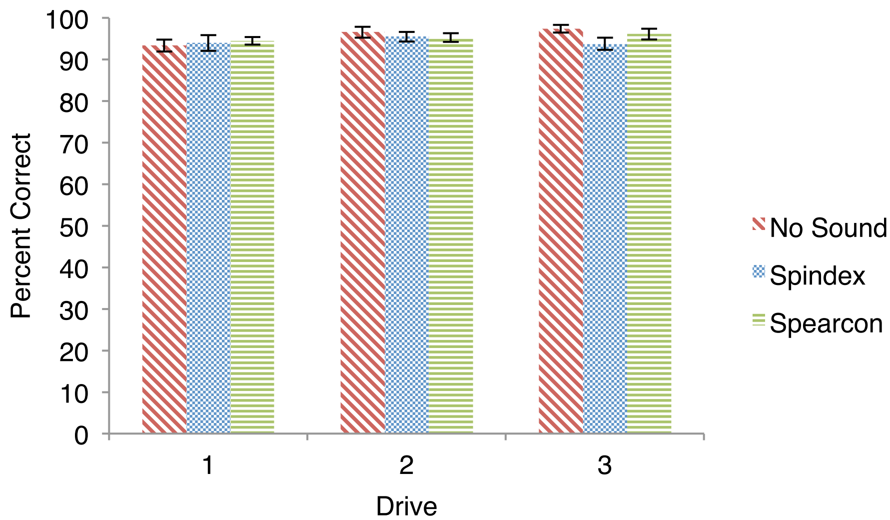


Figure 3.3 - Graph of the mean percent correct selections for each condition across the three drives. Standard error is shown via error bars.

Average Correct Selection Time. For the average time taken for correct selection in the driving blocks the ANOVA revealed a significant main effect of driving block. This was investigated via paired t-tests, which revealed that participants had faster search time in Drive 1 than Drive 3, and in Drive 2 than Drive 3. No significant difference was found for the main effect of condition, nor was there a significant interaction. These analyses can be seen in Table 3.4 and the data are visualized in Figure

3.4 below.

Table 3.4 – Analysis table for the mean time to find a correct song (in seconds) for each condition across the three drives. Significant differences are marked with "*".

Average Correct Selection Time				
ANOVA Results	*F*	df	*p*	*η2*
Main effect of Drive	8.94	1.56, 64.05	.001*	.179
Main effect of Condition	0.89	2.00, 47.00	.419	.036
Interaction	1.66	3.14, 64.05	.183	.075
Post-hoc t-test Results	*t*	df	*p*	
Drive 1 – Drive 2	-1.72	43.00	.092	
Drive 1 – Drive 3	-3.24	43.00	.002*	
Drive 2 – Drive 3	-3.33	43.00	.002*	

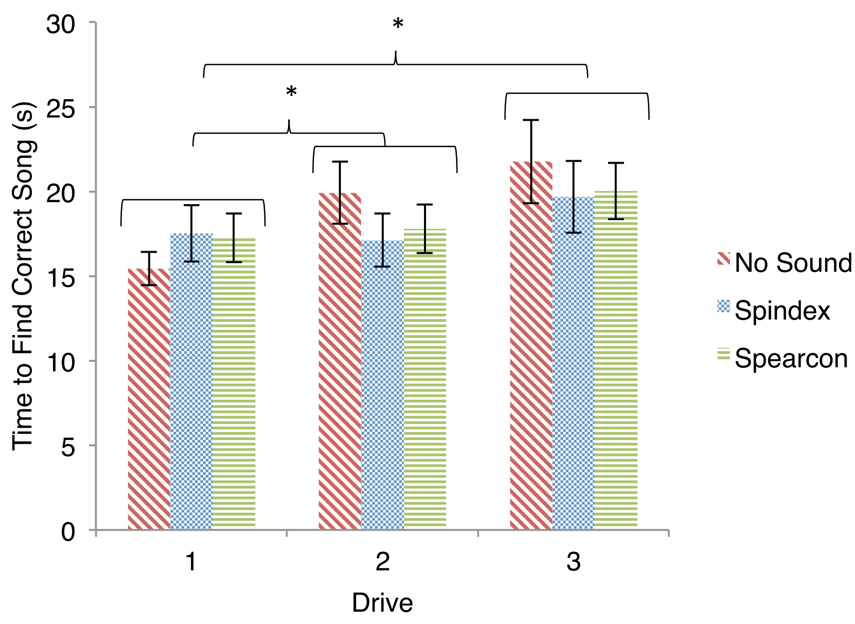


Figure 3.4 - Graph of the mean time to find a correct song (in seconds) for each condition across the three drives. Standard error is shown via error bars and significant differences are marked with "*".

Driving Block List Search Performance Summary. These results show that no significant differences were found in search task performance aside from differences between Blocks 1 and 3 and Blocks 2 and 3 in regards to number of searches and the speed at which participants completed their searches. The trend of these results point to participants actually performing more total searches and doing correct searches more quickly in Drives 1 than 3 and 2 than 3. These results suggest a potential change in abilities or time spent on the secondary task in Drive 3 as compared to Drives 1 and 2 but no differences between conditions in this period or any interactions between driving block and condition.

3.3.2 Driving Performance

Driving performance measures can often be messy due to individual differences in driving abilities and learning effects of using a driving simulator. To determine if this was an issue with the present data the participants' baseline data (as seen in Table J.2 in Appendix J) were analyzed via 2x3 ANOVA (baselines x condition) for the measures to see if there were any differences across sessions or conditions for the baseline drives on Sessions 1 and 4. The results of these analyses can be seen in Table 3.5 below. The analyses revealed that for the measure of mean distance gap there was a significant main effect of session, and that there was a significant main effect of session and condition for standard deviation of distance gap. Finally, there was a significant main effect of condition for standard deviation of lateral lane position.

Table 3.5 – Analysis tables for the driving variables for each condition across the two base drives in Session 1 and 4. Significant differences are marked with "*".

Mean Distance Gap

ANOVA Results	*F*	df	*p*	*η2*

Main effect of Session	24.26	1.00, 43.00	< .001*	.361
Main effect of Condition	1.64	2.00, 43.00	.205	.071
Interaction	0.25	2.00, 43.00	.782	.011

Standard Deviation Distance Gap

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	14.24	1.00, 44.00	< .001*	.244
Main effect of Condition	4.62	2.00, 44.00	.015*	.174
Interaction	0.47	2.00, 44.00	.629	.021

Mean Lateral Lane Deviation

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	0.23	1.00, 46.00	.637	.005
Main effect of Condition	1.84	2.00, 46.00	.170	.074
Interaction	0.12	2.00, 46.00	.887	.005

Standard Deviation Lateral Lane Deviation

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	0.03	1.00, 42.00	.860	.001
Main effect of Condition	3.78	2.00, 42.00	.031*	.153
Interaction	0.28	2.00, 42.00	.785	.013

No post-hocs were done to investigate these main effects as knowing where these differences lie was not necessary to address the issue of individual differences across session or condition. Instead simple percent change (from baseline) difference score values were calculated for the three testing block drives. These were calculated (prior to outlier analysis) by subtracting the value for each performance measure from the corresponding baseline drive value for that session for each participant. This was then divided by the baseline drive value for that session, all of which was then multiplied by 100. For example, mean lateral lane position for Drive 1was changed into percent change mean lateral lane position through the following equation: ((Drive 1 mean lateral lane position – Baseline 1 mean lateral lane position)/ Baseline 1 mean lateral lane position) x

100. This resulted in a percent change difference score (from baseline) for each variable for each of the 3 dual-task driving blocks. The measures of driving performance resulting from these calculations were percent difference of means and standard deviations for distance gap and lateral lane position (longitudinal and lateral control). These were then analyzed and cleaned of outliers in the same fashion as the other data points via 1.5*IQR and then analyzed via 3x3 mixed model ANOVAs as per the planned analyses. These data can be seen in Table J.3 in Appendix J.

Percent Change Of Mean Distance Gap. For percent change in mean distance gap there was a significant main effect of drive, with the post-hocs revealing that participants had significantly higher percent change in average distance gap in Drive 3 than in Drive 1 or Drive 2. The analysis revealed no significant main effect of condition, nor a significant interaction. These analyses can be seen in Table 3.6 and the data are visualized in Figure 3.5 below.

Table 3.6 – Analysis table for the percent change in mean distance gap from baseline for each condition across the three drives. Significant differences are marked with "*".

Percent Change of Mean Distance Gap

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Drive	1.13	2.00, 82.00	< .001*	.228
Main effect of Condition	1.85	2.00, 41.00	.171	.083
Interaction	1.68	4.00, 82.00	.163	.076
Post-hoc t-test Results	*t*	df	*p*	
Drive 1 – Drive 2	0.49	43.00	.623	
Drive 1 – Drive 3	-3.68	43.00	.001*	
Drive 2 – Drive 3	-5.10	43.00	< .001*	

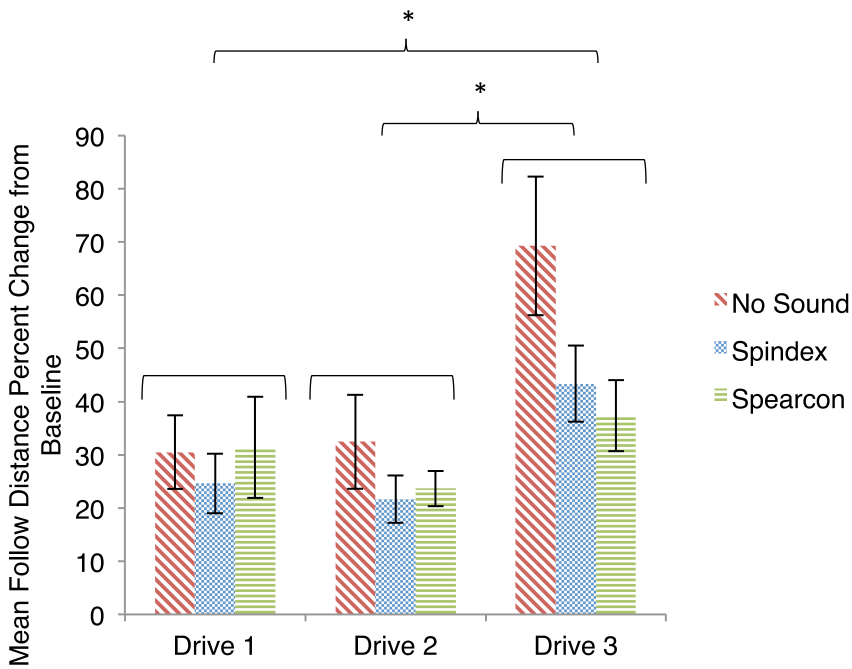


Figure 3.5 - Graph of the percent difference from baseline of mean follow distance gap for each condition across the three drives. Standard error is shown via error bars and significant differences are marked with "*".

Percent Change Of STD Distance Gap. For percent change in standard deviation of distance gap (accordion effect) there was a significant main effect of drive, with the post-hocs seen in the table below revealing that participants had significantly higher percent change in standard deviation of distance gap in Drive 3 than in Drive 1 or Drive 2. The analysis revealed no significant main effect of condition, nor a significant interaction. These analyses can be seen in Table 3.7 and the data are visualized in Figure 3.6 below.

Table 3.7 – Analysis table for the percent change from baseline of the standard deviation of follow distance for each condition across the three drives. Significant differences are marked with "*".

Percent Change of Standard Deviation of Distance Gap

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Drive	64.90	1.30, 51.82	< .001*	.619
Main effect of Condition	1.11	2.00, 40.00	.341	.052
Interaction	0.87	2.59, 51.82	.448	.042

Post-hoc t-test Results	*t*	df	*p*
Drive 1 – Drive 2	0.42	42	.676
Drive 1 – Drive 3	-7.54	42	< .001*
Drive 2 – Drive 3	-10.01	42	< .001*

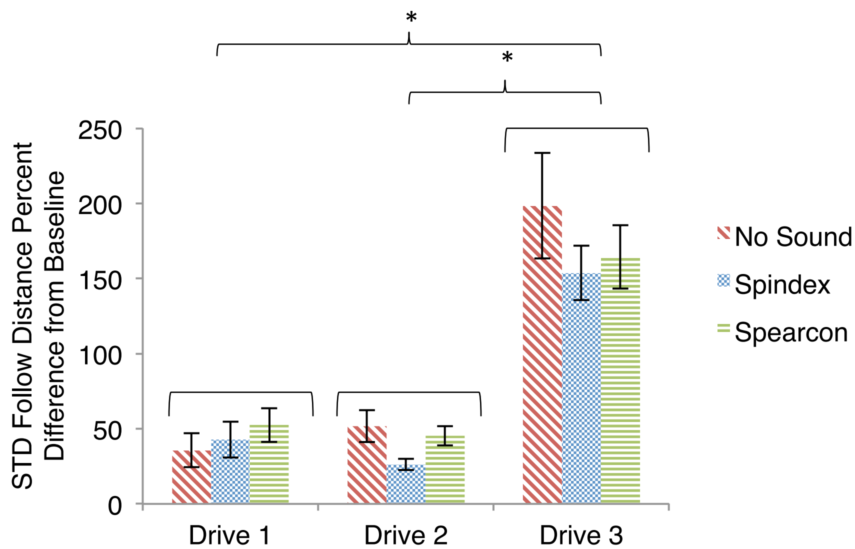


Figure 3.6 - Graph of the percent difference from baseline of the standard deviation of follow distance for each condition across the three drives. Standard error is shown via error bars and significant differences are marked with "*".

Percent Change Of Mean Lateral Lane Position. The analysis of percent change in average lateral lane position revealed a significant main effect of drive, with post-hocs showing that participants had higher average percent change in lateral lane position in Drive 3 than in Drive 1 or Drive 2. The analysis revealed no significant main effect of condition, nor a significant interaction. These analyses can be seen in Table 3.8 and the data are visualized in Figure 3.7 below.

Table 3.8 – Analysis table for the percent change from baseline of the mean lane deviation for each condition across the three drives. Significant differences are marked with "*".

Percent Change of Mean Lateral Lane Position

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Drive	4.66	2.00, 94.00	.012*	.090
Main effect of Condition	0.50	2.00, 47.00	. 609	.021
Interaction	1.98	4.00, 94.00	.103	.078
Post-hoc t-test Results	*t*	df	*p*	
Drive 1 – Drive 2	0.55	49.00	.584	
Drive 1 – Drive 3	-2.05	49.00	.046*	
Drive 2 – Drive 3	-3.33	49.00	.002*	

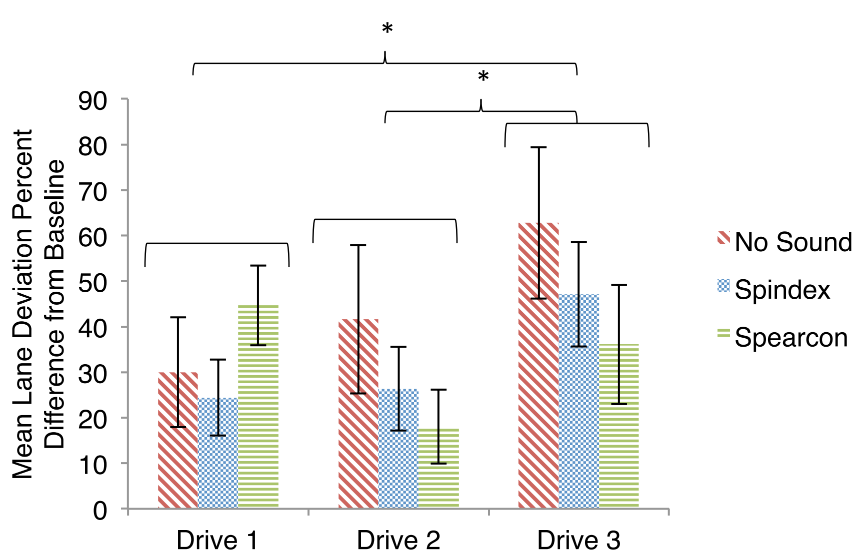


Figure 3.7 - Graph of the percent difference from baseline of mean lane deviation for each condition across the three drives. Standard error is shown via error bars

and significant differences are marked with "*".

Percent Change Of STD Lateral Lane Position. For the percent change in standard deviation of lateral lane position (back and forth between the lines) there was a main effect of drive. Post-hocs revealed that participants had significantly higher percent change in standard deviation of lateral lane position in Drive 3 than in Drive 2. The analysis revealed no significant main effect of condition, nor a significant interaction. These analyses can be seen in Table 3.9 and the data are visualized in Figure 3.8 below.

Table 3.9 – Analysis table for the percent change from baseline of the standard deviation of lane deviation for each condition across the three drives. Significant differences are marked with "*".

Percent Change of Standard Deviation of Distance Gap

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Drive	4.37	1.69, 75.91	.024*	.088
Main effect of Condition	0.46	2.00, 45.00	.635	.020
Interaction	2.07	3.37, 75.91	.104	.084
Post-hoc t-test Results	*t*	df	*p*	
Drive 1 – Drive 2	1.48	47.00	.147	
Drive 1 – Drive 3	-1.10	47.00	.278	
Drive 2 – Drive 3	-4.26	47.00	< .001*	

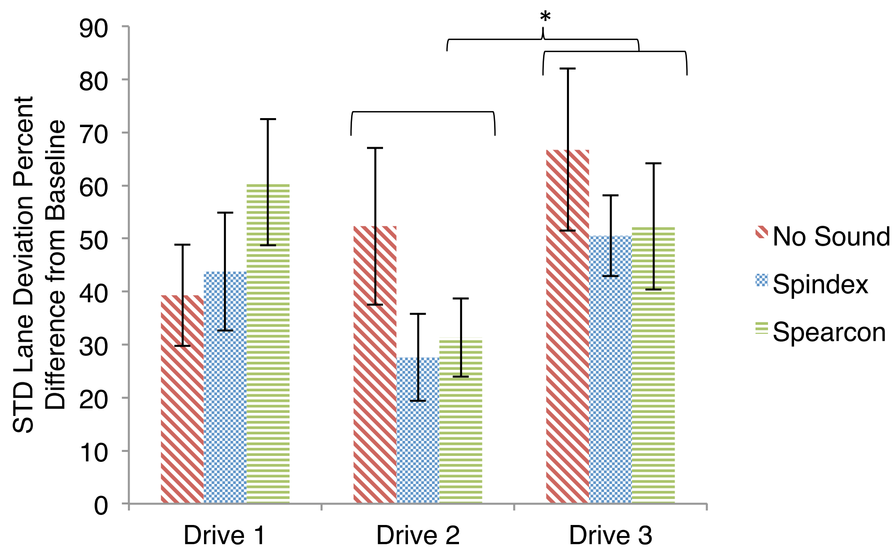


Figure 3.8 - Graph of the percent difference from baseline for standard deviation of lane deviation for each condition across the three drives. Standard error is shown via error bars and significant differences are marked with "*".

Driving Blocks Driving Performance Summary. The results of these analyses suggest that participants had higher following distance, slower reaction time to changes in the lead car, and worse lane keeping performance in Drive 3 than in Drives 1 or 2 across all the conditions. This suggests Drive 3 was more difficult than Drive 1 or 2 as was the aim. However, no main effects of condition or any significant interactions were found, suggesting that there were no differences between the conditions in regards to driving performance.

3.3.3 Driving Block Visual Behaviors

The measures of visual behaviors in the driving simulator were collected for each of the three driving scenarios as well as the driving baselines. The data for these baselines can be seen in Table J.4 in Appendix J. The two baselines were compared across the calculated measures via 2x3 ANOVAs (baseline x condition) to ensure there were no differences between the two baselines or conditions. The results of the ANOVA for time off rate, glance count rate, and average dwell length revealed no significant main effect of baseline, condition, nor an interaction, although as seen in Table 3.10 below some analyses were approaching significance

Table 3.10 - Analysis tables for the baseline comparisons of visual attention for each condition across the two baselines in Session 1 and 4. Significant differences are marked with "*".

Time off Rate

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	0.77	1.00, 43.00	.386	.018
Main effect of Condition	0.62	2.00, 43.00	.542	.028
Interaction	3.02	2.00, 43.00	.059	.123

Glance Count Rate

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	1.36	1.00, 43.00	.250	.031
Main effect of Condition	0.38	2.00, 43.00	.685	.017
Interaction	2.08	2.00, 43.00	.137	.088

Average Dwell Length

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	3.20	1.00, 45.00	.080	.066
Main effect of Condition	2.97	2.00, 45.00	.061	.117
Interaction	2.30	2.00, 45.00	.112	.093

Although these analyses revealed no significant main effects, there were a number of values approaching significance, and therefore it was decided to create difference scores to ensure equal comparisons were being made across the analyses. To do this for the three driving scenarios, difference scores were calculated based on the base drive eye-tracking data. These were simply subtractive measures instead of percent change difference scores, as some data value were equal to 0 in the baseline drives and therefore would not be usable to create percentage change difference scores. The three measures used from the raw data were time off rate (percent time off the driving task), glance count rate (average number of glances off the road per minute), and average dwell length (average length of time per dwell off the road). This meant the calculated measures were change in time off rate (driving block time off rate – baseline time off rate), change in glance count rate, and change in average dwell length, the data for which can be seen in Table J.5 in Appendix J.

Change in Time Off Rate. For time off rate the analysis revealed a significant main effect of drive. Post-hoc analyses showed that this difference was due to significantly higher percent of time off the road for Drive 1 than Drives 2 or 3, and more percent time off for Drive 2 than Drive 3. For the main effect of condition there was also a significant difference. The t-test post-hoc analyses revealed that participants in the No-Sound condition had more percent time eyes off the road across the drives than participants in the Spindex condition or the Spearcon condition, and that participants in the Spearcon condition had a higher rate of eyes off the road than the participants in the Spindex condition. No significant interaction was found. These analyses can be seen in Table 3.11 and the data are visualized in Figure 3.9 below.

Table 3.11 - Analysis table for the change in time off rate for each condition across the three drives. Significant differences are marked with "*".

Change in Time Off Rate

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Drive	37.97	1.35, 56.50	< .001*	.475
Main effect of Condition	17.48	2.00, 42.00	< .001*	.454

Interaction	1.41	2.69, 56.50	.251	.063
Post-hoc t-test Results	*t*	df	*p*	
Drive 1 – Drive 2	5.31	44.00	< .001*	
Drive 1 – Drive 3	6.91	44.00	< .001*	
Drive 2 – Drive 3	4.35	44.00	< .001*	
No-Sound - Spindex	7.86	27.00	< .001*	
No-Sound - Spearcon	3.23	29.00	.003*	
Spindex - Spearcon	-2.25	28.00	.033*	

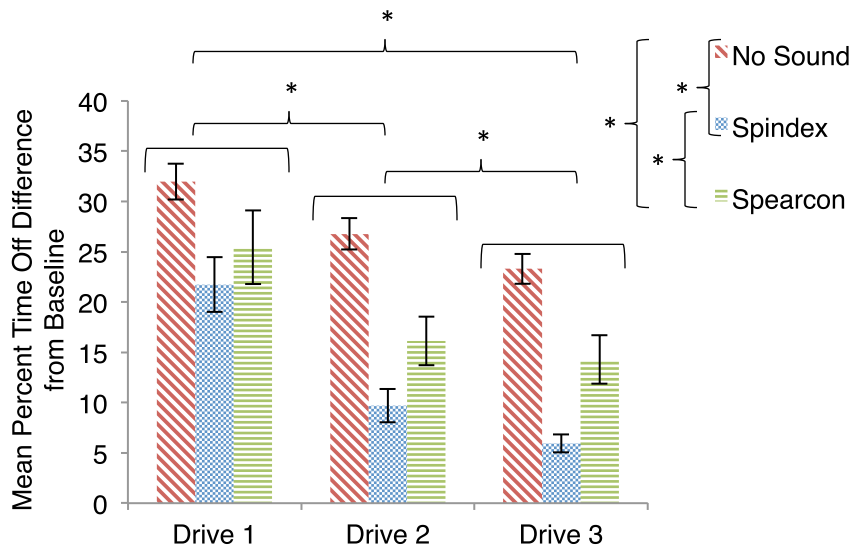


Figure 3.9 - Graph of the difference score from baseline for mean time off rate (percent) for each condition across the three drives. Standard error is shown via error bars and significant differences are marked with "*".

Change in Glance Count Rate. In the analysis of glance count rate a significant main effect of drive was found. The follow-up test revealed that when performing Drive 1 participants had a significantly higher rate of glances off the road than in Drive 2 or Drive 3. In addition participants had a higher rate of glances off the road in Drive 2 than Drive 3. There was also a significant main effect of condition, with the post-hocs revealing that the participants in the No-Sound condition had a higher glance count rate than those in the Spindex condition, and was approaching significance for those in the Spearcon condition. No significant interaction was found in the analysis. These analyses can be seen in Table 3.12 and the data are visualized in Figure 3.10 below.

Table 3.12 - Analysis table for the change in glance count rate for each condition across the three drives. Significant differences are marked with "*".

Change in Glance Count Rate				
ANOVA Results	*F*	df	*p*	*η2*
Main effect of Drive	49.23	1.73, 72.50	< .001*	.540
Main effect of Condition	4.41	2.00, 42.00	.018*	.174
Interaction	0.81	3.45, 72.50	.509	.037
Post-hoc t-test Results	*t*	df	*p*	
Drive 1 – Drive 2	6.53	44.00	< .001*	
Drive 1 – Drive 3	8.48	44.00	< .001*	
Drive 2 – Drive 3	3.60	44.00	.001*	
No-Sound - Spindex	3.73	27.00	.001*	
No-Sound - Spearcon	-2.25	28.00	.063	
Spindex - Spearcon	-0.82	29	.426	

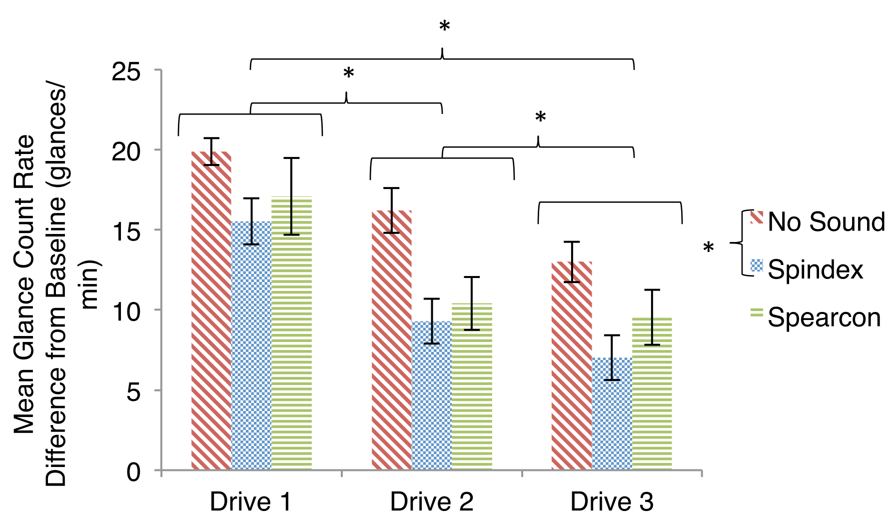


Figure 3.10 - Graph of the difference score from baseline for mean glance rate

(glances/minute) for each condition across the three drives. Standard error is shown via error bars and significant differences are marked with "*".

Change in Average Dwell Length. For average dwell length there was no significant main effect of drive, condition, nor any interaction. These analyses can be seen in Table 3.13 and the data are visualized in Figure 3.11 below.

Table 3.13 - Analysis table for the change in average dwell length for each condition across the three drives.

Change in Average Dwell Length

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Drive	0.77	1.22, 53.61	.407	.017
Main effect of Condition	0.01	2.00, 44.00	.992	< .001
Interaction	0.75	2.44, 53.61	.502	.033

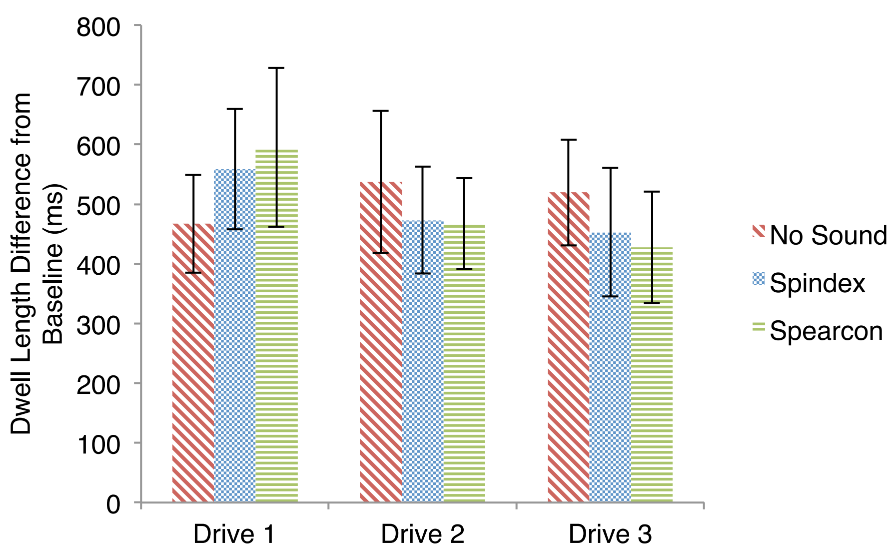


Figure 3.11 - Graph of the difference score from baseline for mean dwell length (ms) for each condition across the three drives. Standard error is shown via error bars.

Driving Blocks Eye-tracking Summary. These results point to a higher percentage of time eyes off the road, higher and higher glance count rate in Drive 1 than in Drives 2 or 3, and in Drive 2 than in Drive 3. This suggests participants decreased their eyes off the road time with practice in the dual task scenario and when the drive became more difficult they adjusted their approach as well. For the conditions the data shows that across the three drives participants in the No-Sound condition had higher percent time off the road and a higher glance count rate off the road than either participants in either the Spindex or Spearcon conditions. Further, those in the Spearcon condition had a higher percent time eyes off the road than those in the Spindex condition. The lack of any differences in average dwell length suggest that participants did not have longer glances off the road but simply more of them in these block or condition differences.

3.3.4 Driving Block Heart Rate

For heart rate measures, individual differences across sessions and participants can create difficulties with comparing directly. For this reason two 4x3 (baseline x condition) mixed model ANOVAs (with Huynh-Feldt corrections) were performed across the two HR measures collected (as seen in Table J.6 in Appendix J) to check for any baseline differences. As seen in Table 3.12 below, the analyses revealed that for mean HR there was a significant main effect of session, a significant main effect of condition, but no significant interaction. For mean HRV the analysis revealed a significant main effect of session, but no significant main effect of condition, nor a significant interaction.

Table 3.14 - Analysis table for the comparisons of HR and HRV baselines for the three conditions across the three sessions. Significant differences are marked with "*".

Mean Heart Rate

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	2.79	2.72, 122.59	.049*	.058
Main effect of Condition	4.28	2.00, 45.00	.020*	.160
Interaction	1.30	5.45, 122.59	.266	.055

Mean Heart Rate Variance

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	3.03	3.00, 132.00	.032*	.064
Main effect of Condition	1.22	3.00, 44.00	.306	.052
Interaction	1.17	6.00, 132.00	.325	.051

These results show that difference scores were needed to control for these individual differences and so percent-change (from baseline) difference scores were calculated mean (HR) and mean (HRV). These scores were calculated in the same way as the driving variables and were done for both measures across all 3 driving blocks, the descriptive data for which can be seen in Table J.7 in Appendix J.

Percent Change in Mean Heart Rate. For the percentage change in mean HR there was a significant main effect of drive. Post-hoc analyses revealed the differences to be that participants in Drive 1 had a higher percent change from baseline than when performing Drive 2 or Drive 3. The analysis revealed no significant main effect of condition, nor a significant interaction. These analyses can be seen in Table 3.15 and the data are visualized in Figure 3.12 below.

Table 3.15 - Analysis table for the percent change in mean heart rate for each condition across the three drives. Significant differences are marked with "*".

Percent Change in Mean Heart Rate				
ANOVA Results	*F*	df	*p*	*η2*
Main effect of Drive	30.75	1.43, 63.05	< .001*	.411
Main effect of Condition	0.29	2.00, 44.00	.748	.013
Interaction	1.98	2.87, 63.05	.129	.082
Post-hoc t-test Results	*t*	df	*p*	
Drive 1 – Drive 2	6.22	46.00	< .001*	
Drive 1 – Drive 3	5.30	46.00	< .001*	
Drive 2 – Drive 3	-0.33	46.00	.741	

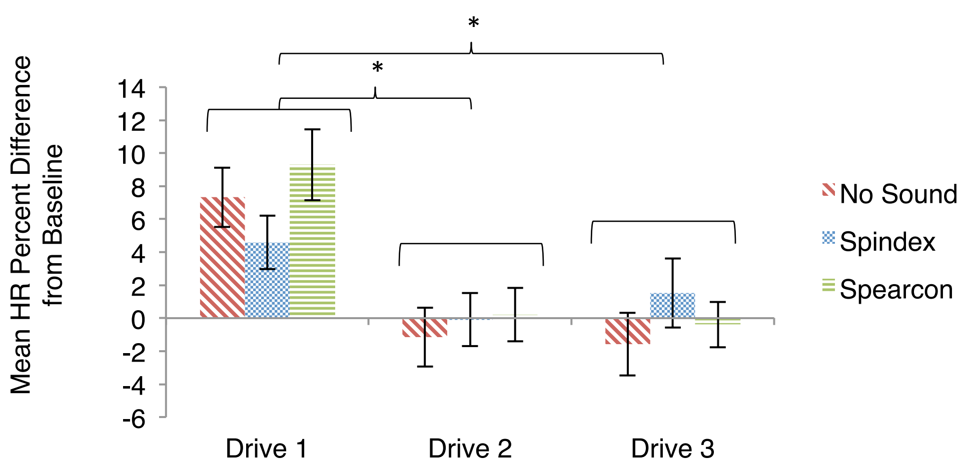


Figure 3.12 - Graph of the percent difference from baseline of mean heart rate for each condition across the three drives. Standard error is shown via error bars and significant differences are marked with "*".

Percent Change in Mean Heart Rate Variability. The analysis of percent change for mean HRV revealed no significant main effect of drive, condition, nor a significant interaction. These analyses can be seen in Table 3.16 and the data are visualized in Figure 3.13 below.

Table 3.16 - Analysis table for the percent change in mean heart rate variance for each condition across the three drives. Significant differences are marked with "*".

Percent Change in Mean Heart Rate Variance

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Drive	3.29	1.42, 55.41	.060	.078
Main effect of Condition	2.71	2.00, 39.00	.079	.122
Interaction	1.29	2.84, 55.41	.287	.062

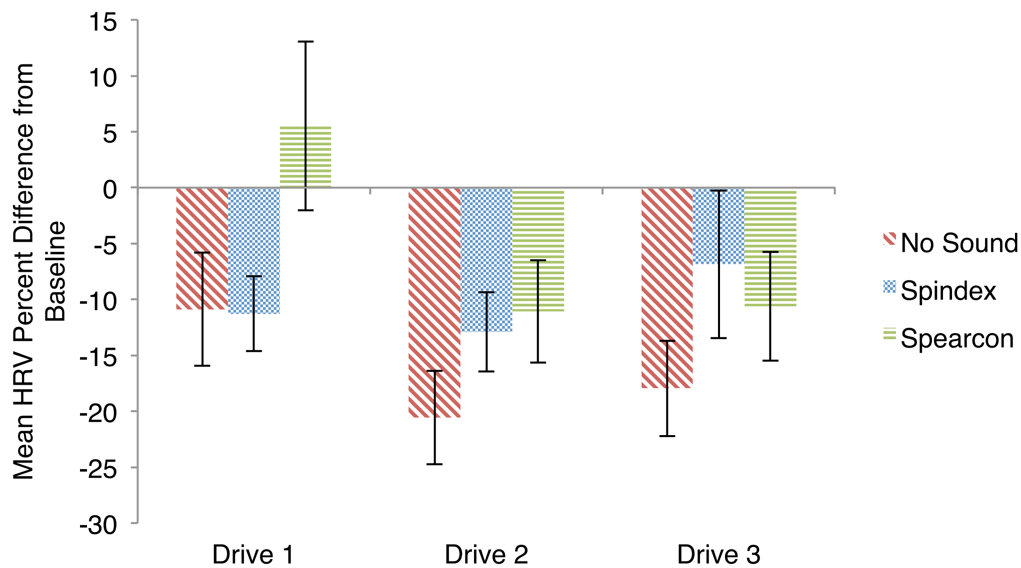


Figure 3.13 - Graph of the percent difference from baseline of mean heart rate variance for each condition across the three drives. Standard error is shown via error bars.

Driving Blocks Objective Workload Summary. The results of the heart rate analyses gave mixed results. While mean HR showed that participants had significantly higher percent change in heart rate in Drive 1 than Drives 2 or 3 from their baselines, there were no differences in mean HRV, although it was approaching significance for the main effect of drive. For the main effect of condition, neither measures were significant, showing there was not a reliable difference in physiological response via HR or HRV between the conditions.

3.3.5 Driving Block Subjective Workload

The NASA-TLX measure of subjective workload was analyzed by 3x3 mixed models drive by condition ANOVAs for each subscale and the composite score. The descriptive data for these measures can be seen in Table J.8 in Appendix J and below are the results of these analyses.

Mental Workload. For the subscale of mental workload participants' ratings showed a significant main effect of drive. Post-hocs revealed that participants rated the first drive as having significantly higher mental demand than Drive 2 or 3. There was no significant main effect of condition, nor a significant interaction found. These analyses can be seen in Table 3.17 and the data are visualized in Figure 3.14 below.

Table 3.17 - Analysis table for subjective mental workload NASA-TLX scores for each condition across the three drives. Significant differences are marked with "*".

Mental Workload

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Drive	4.93	1.47, 69.13	.018*	.095
Main effect of Condition	0.62	1.00, 47.00	.545	.026
Interaction	0.41	2.94, 69.13	.743	.017
Post-hoc t-test Results	*t*	df	*p*	
Drive 1 – Drive 2	2.54	49.00	.014*	
Drive 1 – Drive 3	2.14	49.00	.038*	
Drive 2 – Drive 3	-0.93	49.00	.355	

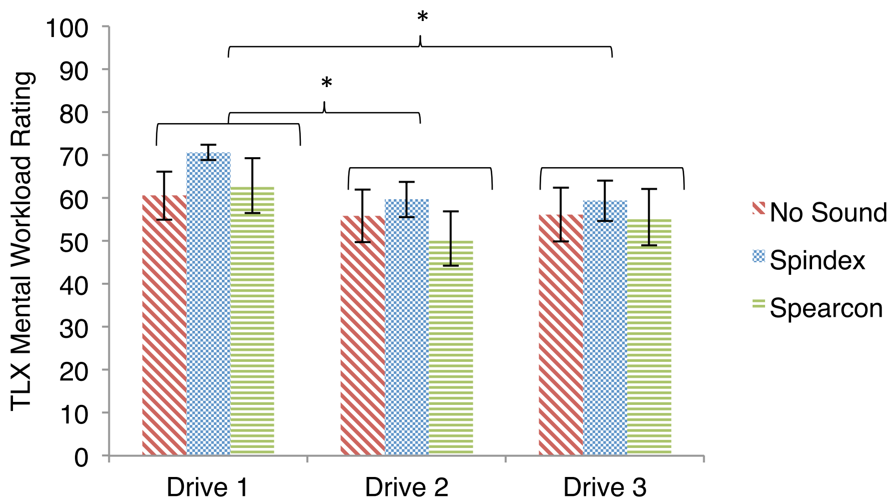


Figure 3.14 - Graph of the NASA-TLX ratings for mental workload for each condition across the three drives. Standard error is shown via error bars and significant differences are marked with "*".

Physical Workload. For physical workload participants ratings revealed a significant main effect of drive, with post-hocs revealing that participants rated the first drive as having significantly lower physical demand than Drive 3. The condition main effect was also found to be significant, with post-hocs revealing that the Spearcon condition was rated as having higher physical workload than the Spindex condition, and approaching significance for the No-Sound condition. No significant interaction was found for physical workload. These analyses can be seen in Table 3.18 and the data are visualized in Figure 3.15 below.

Table 3.18 - Analysis table for subjective physical workload NASA-TLX scores for each condition across the three drives. Significant differences are marked with "*".

Physical Workload

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Drive	3.56	1.51, 70.93	.046*	.070
Main effect of Condition	3.39	2.00, 47.00	.042*	.126

Interaction	0.23	3.02, 70.93	.873	.010
Post-hoc t-test Results	*t*	df	*p*	
Drive 1 – Drive 2	-1.84	49.00	.071	
Drive 1 – Drive 3	-2.22	49.00	.031*	
Drive 2 – Drive 3	-0.73	49.00	.467	
No-Sound - Spindex	0.28	31.00	.781	
No-Sound - Spearcon	1.90	32.00	.067	
Spindex - Spearcon	2.57	31.00	.015*	

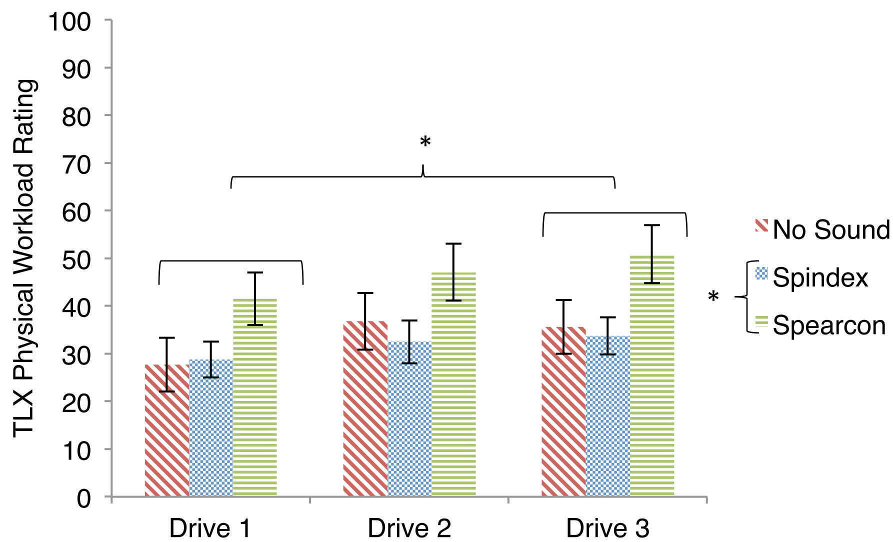


Figure 3.15 - Graph of the NASA-TLX ratings for physical workload for each condition across the three drives. Standard error is shown via error bars and significant differences are marked with "*".

Subjective Performance Rating. For subjective performance, participant ratings showed a significant main effect of drive, with post-hocs revealing higher reported performance on the first drive than Drives 2 or 3. The main effect of condition was also found to be significant, with post-hocs revealing that participants gave higher performance ratings in the No Sound condition than the Spindex condition, or the

Spearcon condition, and the Spindex condition participants giving higher ratings than those in the Spearcon condition. No significant interaction was found for subjective performance. These analyses can be seen in Table 3.19 and the data are visualized in Figure 3.16 below.

Table 3.19 - Analysis table for subjective performance NASA-TLX scores for each condition across the three drives. Significant differences are marked with "*".

Performance				
ANOVA Results	*F*	df	*p*	*η2*
Main effect of Drive	20.85	1.96, 80.40	< .001*	.337
Main effect of Condition	11.75	2.00, 41.00	< .001*	.364
Interaction	0.23	3.02, 70.93	.873	.010
Post-hoc t-test Results	*t*	df	*p*	
Drive 1 – Drive 2	5.56	43.00	< .001*	
Drive 1 – Drive 3	4.95	43.00	< .001*	
Drive 2 – Drive 3	-0.72	43.00	.476	
No-Sound - Spindex	2.42	29.00	.022*	
No-Sound - Spearcon	4.29	27.00	< .001*	
Spindex - Spearcon	2.94	26.00	.007*	

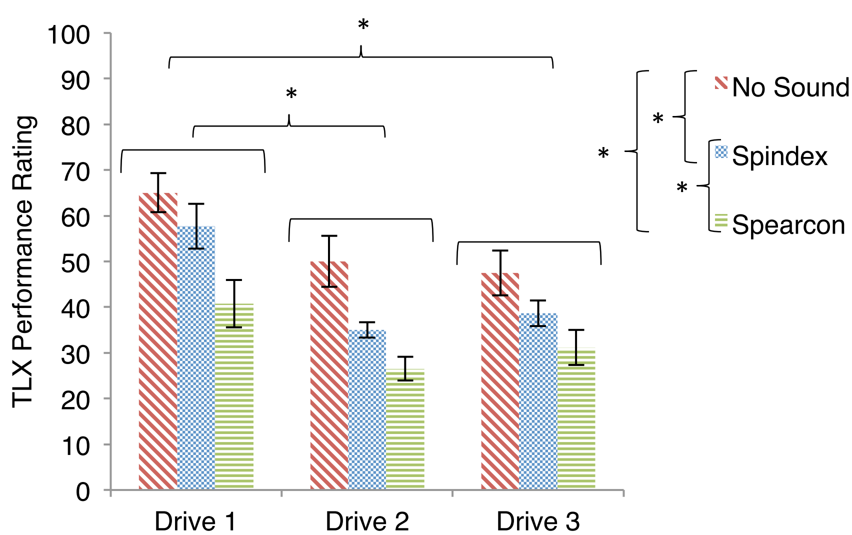


Figure 3.16 - Graph of the NASA-TLX ratings for performance for each condition

across the three drives. Standard error is shown via error bars and significant differences are marked with "*".

Subjective Effort. In regards to subjective effort, the ratings from participants revealed a significant main effect of drive, with post-hocs revealing higher subjective effort in the first drive than Drive 2 or Drive 3. The main effect of condition was not significant, nor was any interaction. These analyses can be seen in Table 3.20 and the data are visualized in Figure 3.17 below.

Table 3.20 - Analysis table for subjective effort NASA-TLX scores for each condition across the three drives. Significant differences are marked with "*".

Effort				
ANOVA Results	*F*	df	*p*	*η2*
Main effect of Drive	9.85	1.78, 81.82	< .001*	.176
Main effect of Condition	2.13	2.00, 46.00	.130	.085
Interaction	0.17	3.56, 81.82	.923	.007
Post-hoc t-test Results	*t*	df	*p*	
Drive 1 – Drive 2	4.06	48.00	< .001*	
Drive 1 – Drive 3	2.72	48.00	.009*	
Drive 2 – Drive 3	-1.76	48.00	.084	

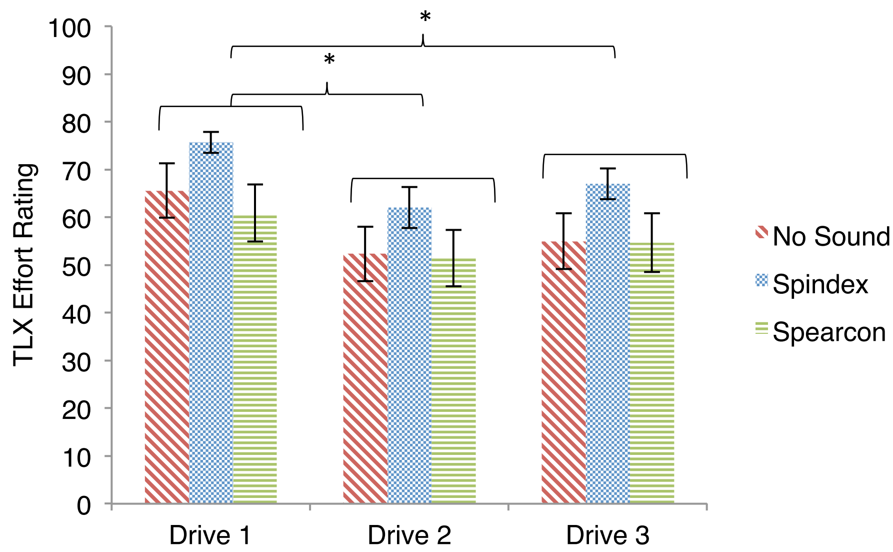


Figure 3.17 - Graph of the NASA-TLX ratings for effort for each condition across the three drives. Standard error is shown via error bars and significant differences are marked with "*".

Temporal Workload. For perceived temporal workload the analysis revealed a significant main effect of drive, with post-hocs showing that participants rated Drive 2 as having lower temporal workload than Drive 1 or Drive 3. The analysis showed no significant main effect of condition, nor any significant interaction. These analyses can be seen in Table 3.21 and the data are visualized in Figure 3.18 below.

Table 3.21 - Analysis table for subjective temporal workload NASA-TLX scores for each condition across the three drives. Significant differences are marked with "*".

Temporal Workload				
ANOVA Results	*F*	df	*p*	*η2*
Main effect of Drive	5.46	1.68, 80.66	.009*	.102
Main effect of Condition	2.34	2.00, 48.00	.108	.089
Interaction	2.22	3.36, 80.66	.085	.085
Post-hoc t-test Results	*t*	df	*p*	
Drive 1 – Drive 2	2.74	50.00	.008*	

Drive 1 – Drive 3	1.67	50.00	.102
Drive 2 – Drive 3	-2.10	50.00	.041*

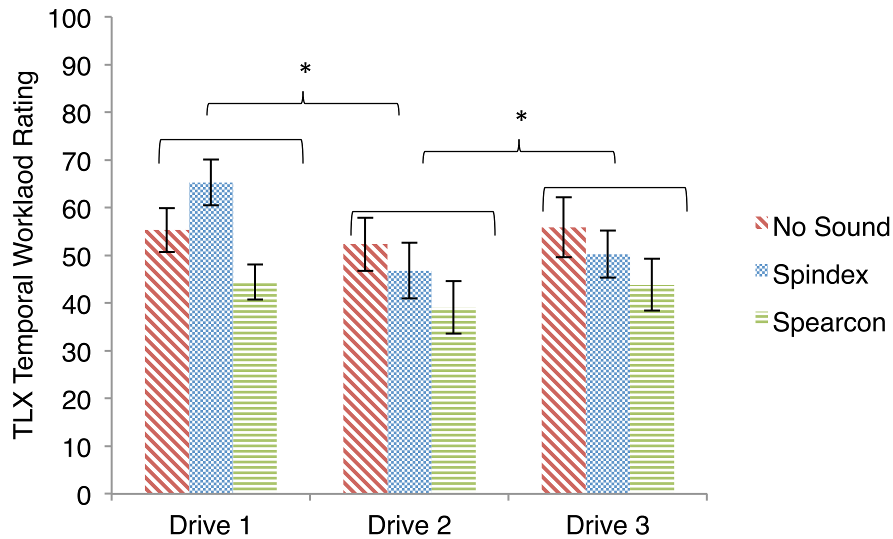


Figure 3.18 - Graph of the NASA-TLX ratings for temporal workload for each condition across the three drives. Standard error is shown via error bars and significant differences are marked with "*".

Frustration. For subjective ratings of frustration the analysis revealed a significant main effect of drive, with post-hocs showing that participants rated Drive 1 as having higher frustration than Drive 2 or Drive 3. The analysis showed no significant main effect of condition, nor any significant interaction. These analyses can be seen in Table 3.22 and the data are visualized in Figure 3.19 below.

Table 3.22 - Analysis table for subjective frustration NASA-TLX scores for each condition across the three drives. Significant differences are marked with "*".

Frustration				
ANOVA Results	*F*	df	*p*	*η2*
Main effect of Drive	12.64	1.54, 73.84	< .001*	.208
Main effect of Condition	0.11	2.00, 48.00	.900	.004
Interaction	1.23	3.08, 73.84	.305	.049

Post-hoc t-test Results	*t*	df	*p*
Drive 1 – Drive 2	4.00	50.00	< .001*
Drive 1 – Drive 3	3.53	50.00	.001*
Drive 2 – Drive 3	-1.37	50.00	.177

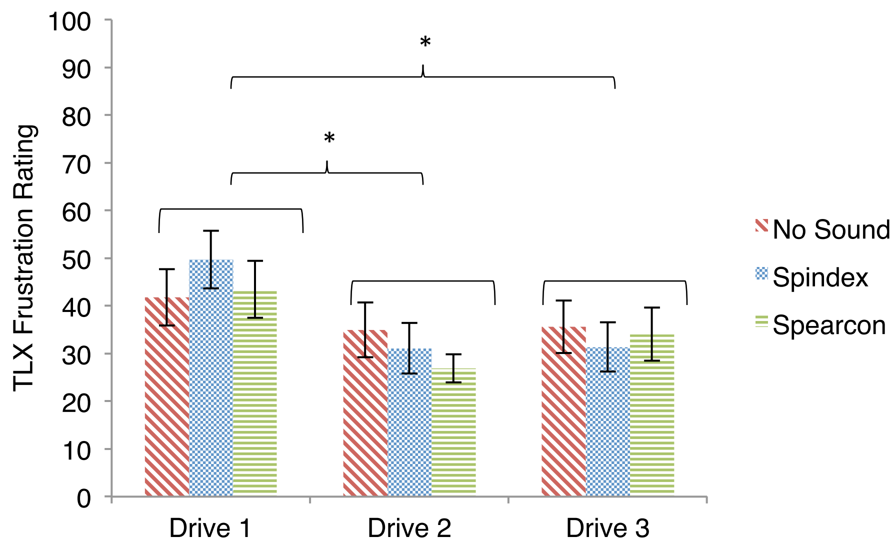


Figure 3.19 - Graph of the NASA-TLX ratings for frustration for each condition across the three drives. Standard error is shown via error bars and significant differences are marked with "*".

NASA-TLX Composite Score. In regards to the NASA-TLX composite score of subjective workload the analysis revealed a significant main effect of drive, with post-hocs showing that participants rated Drive 1 as having higher workload than Drive 2 or Drive 3, and Drive 3 having higher total workload than Drive 2. The analysis showed no significant main effect of condition, nor any significant interaction. These analyses can be seen in Table 3.23 and the data are visualized in Figure 3.20 below.

Table 3.23 - Analysis table for subjective composite workload NASA-TLX scores for each condition across the three drives. Significant differences are marked with "*".

Composite Workload

ANOVA Results	F	df	p	$\eta 2$
Main effect of Drive	9.81	1.41, 67.52	.001*	.170
Main effect of Condition	0.66	2.00, 48.00	.570	.027
Interaction	1.23	2.81, 67.52	.305	.049

Post-hoc t-test Results	t	df	p
Drive 1 – Drive 2	3.77	50.00	< .001*
Drive 1 – Drive 3	2.64	50.00	.011*
Drive 2 – Drive 3	-2.24	50.00	.030*

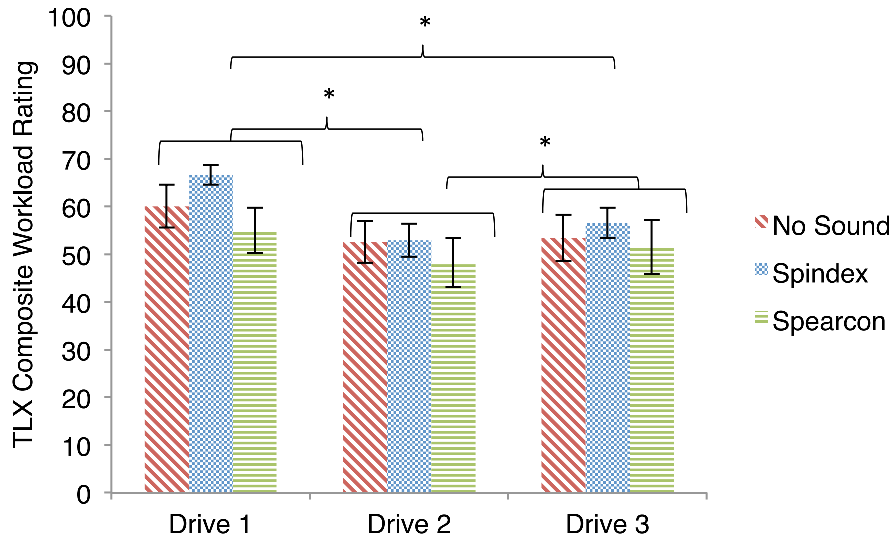


Figure 3.20 - Graph of the NASA-TLX ratings for composite workload for each condition across the three drives. Standard error is shown via error bars and significant differences are marked with "*".

Driving Blocks Subjective Workload Summary. These results show a trend of workload across the three drives with the first drive as having more effort, mental demand, higher frustration, and higher total workload in Drive 1 than Drives 2 or 3. This means that the first drive was more difficult for participants than the other two drives, in regards to these factors, suggesting that as participants practiced they decreased their workload when completing this type of dual task. Similarly, the result of lower subjective ratings of temporal workload in Drive 2 than Drive 1 or 3 suggest that practice in the dual task situation decreased the temporal demand on participants in Drive 2, which is the same type of scenario as in Drive 1, but that putting participants in a more difficult driving scenario (i.e., in Drive 3) increased this temporal load again. In addition, participants reported having higher total workload in Drive 3 than Drive 2, suggesting the curvy scenario of Drive 3 made the dual task more difficult. The finding of the first drive having less physical demand than Drive 3 also supports this increase in workload in Drive 3, potentially due to the extra steering and attention that was demanded to the road in the curvy drive.

The results of the main effect of condition reveal that participants found the Spearcon condition to have higher physical workload than the No-Sound or the Spindex condition. At the same time participants subjectively reported higher performance with the No-Sound condition than the Spearcon or Spindex conditions. Finally, participants in the Spindex condition also reported having higher performance than those in the Spearcon condition.

3.4 Training Block Results

As stated earlier, the following section investigates the training block results from the standpoint of the average across blocks per session. The descriptive data for the results per block can be seen alongside the average per session in Appendix J and the graphs of the data per block can be seen in Appendix K.

3.4.1 Training Blocks List Search Performance

For measures of cell phone list search task performance during the training blocks the same four measures were used as in the driving blocks including number of trials, number of songs correctly selected, percent accuracy, and average time to find a correct song. The descriptive data for these measures is in Appendix J, shown by block in Table J.9 and shown by session average in Table J.10.

Number of Searches. For the number of trials in the training blocks the ANOVA revealed a significant main effect of session. Post-hocs for the main effect of condition revealed that participants had a significantly higher mean number of searches per block in Sessions 2, 3, and 4 than in Session 1, and higher number of searches in Session 4 than in Session 2. No significant difference was found for the main effect of condition, nor was a significant interaction found. These analyses can be seen in Table 3.24 and the data are visualized in Figure 3.21 below.

Table 3.24 - Analysis table for mean number of searches for each condition across the mean of the training blocks for the four sessions. Significant differences are marked with "*".

Number of Searches

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	17.52	2.50, 114.93	< .001*	.276
Main effect of Condition	1.48	2.00, 46.00	.238	.061
Interaction	1.09	5.00, 114.93	.371	.045
Post-hoc t-test Results	*t*	df	*p*	
Session 1 – Session 2	-3.86	48.00	< .001*	
Session 1 – Session 3	-4.46	48.00	< .001*	
Session 1 – Session 4	-6.32	48.00	< .001*	
Session 2 – Session 3	-2.30	48.00	.026	
Session 2 – Session 4	-3.15	48.00	.003*	
Session 3 – Session 4	-2.24	48.00	.030	

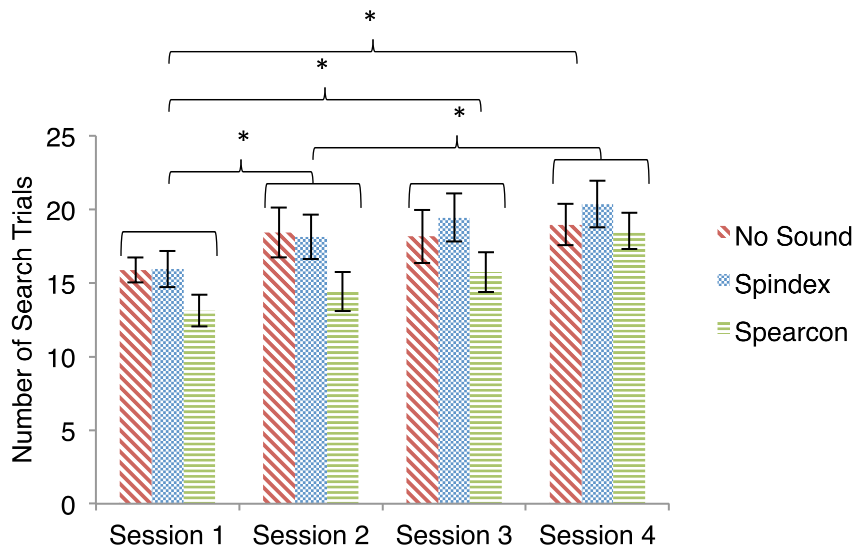


Figure 3.21 - Graph of the mean number of songs searched for each condition across the training block averages for the 4 sessions. Standard error is shown via error bars and significant differences are marked with "*".

Percent Correct. In the analyses for the percent correct selections in the training blocks the data showed a significant main effect of session. Follow up t-tests showed that participants had significantly higher mean percent correct selections in Sessions 2, 3, and 4 than in Session 1. The ANOVA revealed no significant difference in condition, nor any significant interaction. These analyses can be seen in Table 3.25 and the data are visualized in Figure 3.22 below.

Table 3.25 - Analysis table for mean percent correct selections for each condition across the mean of the training blocks for the four sessions. Significant differences are marked with "*".

Percent Correct

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	15.85	1.82, 78.09	< .001*	.269
Main effect of Condition	0.64	2.00, 43.00	.535	.029

Interaction	1.58	3.63, 78.09	.194	.068
Post-hoc t-test Results	*t*	df	*p*	
Session 1 – Session 2	-3.58	45.00	.001*	
Session 1 – Session 3	-4.30	45.00	< .001*	
Session 1 – Session 4	-5.10	45.00	< .001*	
Session 2 – Session 3	-1.09	45.00	.280	
Session 2 – Session 4	-2.38	45.00	.022	
Session 3 – Session 4	-1.44	45.00	.158	

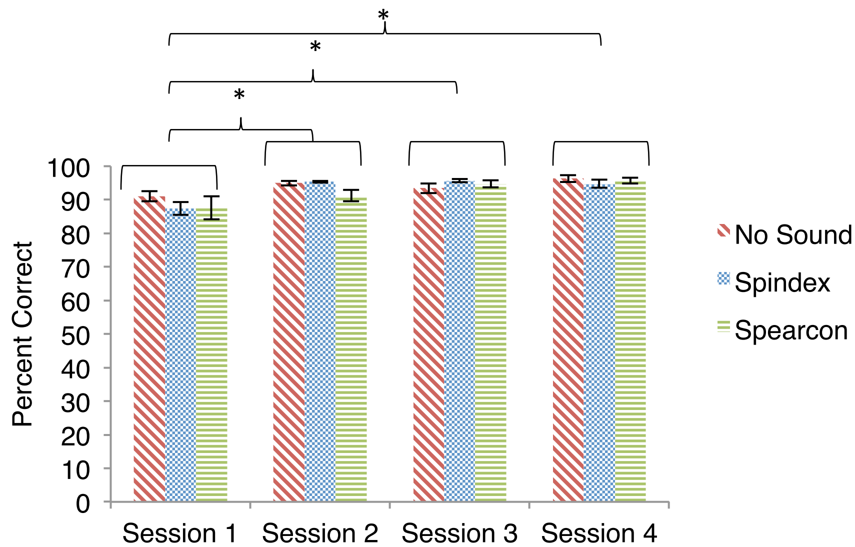


Figure 3.22 - Graph of the mean song selection accuracy (percentage) for each condition across the training block averages for the 4 sessions. Standard error is shown via error bars and significant differences are marked with "*".

Average Correct Selection Time. For the average time taken for correct selections in the training blocks there was a significant main effect of block. Follow up post-hocs showed that participants had a significantly lower block average correct selection time in Sessions 3 and 4 than in Session 1, and lower selection time in Session 4 than Session 2. The ANOVA revealed no significant difference in condition, nor any significant interaction. These analyses can be seen in Table 3.26 and the data are visualized in Figure 3.23 below.

Table 3.26 - Analysis table for mean correct selection time for each condition across the mean of the training blocks for the four sessions. Significant differences are marked with "*".

Average Correct Selection Time				
ANOVA Results	*F*	df	*p*	$\eta 2$
Main effect of Session	10.76	2.64, 116.31	< .001*	.196
Main effect of Condition	0.47	2.00, 44.00	.629	.021
Interaction	1.06	5.29, 116.31	.386	.046
Post-hoc t-test Results	*t*	df	*p*	
Session 1 – Session 2	2.61	46.00	.012	
Session 1 – Session 3	2.99	46.00	.005*	
Session 1 – Session 4	4.58	46.00	< .001*	
Session 2 – Session 3	1.39	46.00	.170	
Session 2 – Session 4	3.11	46.00	.003*	
Session 3 – Session 4	2.46	46.00	.018	

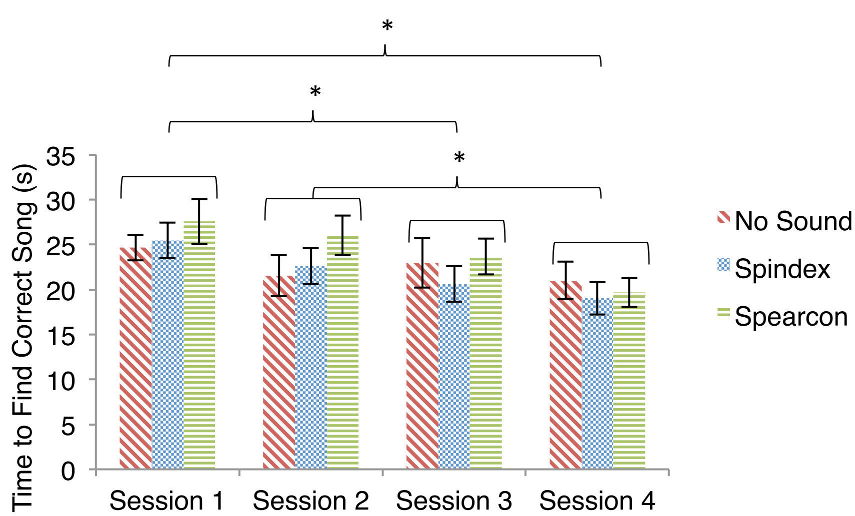


Figure 3.23 - Graph of the mean time to find a correct song for each condition across the training block averages for the 4 sessions. Standard error is shown via error bars and significant differences are marked with "*".

Training Blocks List Search Summary. The results seen in the data show a trend of a higher number of searches and higher percent accuracy of those searches in the last three sessions than the first one. In addition the results of correct selection time show a trend of decreasing time necessary to make a correct selection, which would be expected based on the increasing number of searches and correct songs selected. The lack of any differences between condition suggest participants had no more difficulty in completing the tasks between the conditions.

3.4.2 Training Blocks Ball Drop Performance

For data regarding the participants' performance on the ball drop game during the training blocks two values were pulled, accuracy and balls released. The accuracy measure represents the participants' performance for each block via the percentage of balls they caught as compared to the number released by the game in each block. The measure of balls released from above per block (or the speed of the balls being released) represents the number of balls participants had the opportunity to catch during each block, and is what is determined during the calibration setup at the beginning of each session.

Accuracy. For accuracy of ball catching the ANOVA revealed a significant main effect of block average per session. The post-hoc t-tests revealed that participants had a higher mean percent caught per block in Session 4 than in Sessions 1 or 2. There was no significant difference in regards to condition, nor any significant interaction. These analyses can be seen in Table 3.27 and the data are visualized in Figure 3.24 below.

Table 3.27 - Analysis table for mean ball drop accuracy for each condition across the mean of the training blocks for the four sessions. Significant differences are marked with "*".

Accuracy

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	5.79	3.00, 144.00	.001*	.108
Main effect of Condition	1.07	2.00, 48.00	.351	.043
Interaction	0.42	6.00, 144.00	.863	.017
Post-hoc t-test Results	*t*	df	*p*	
Session 1 – Session 2	-1.75	50.00	.087	
Session 1 – Session 3	-2.37	50.00	.022	
Session 1 – Session 4	-3.92	50.00	< .001*	
Session 2 – Session 3	-.88	50.00	.386	
Session 2 – Session 4	-2.81	50.00	.007*	
Session 3 – Session 4	-1.56	50.00	.126	

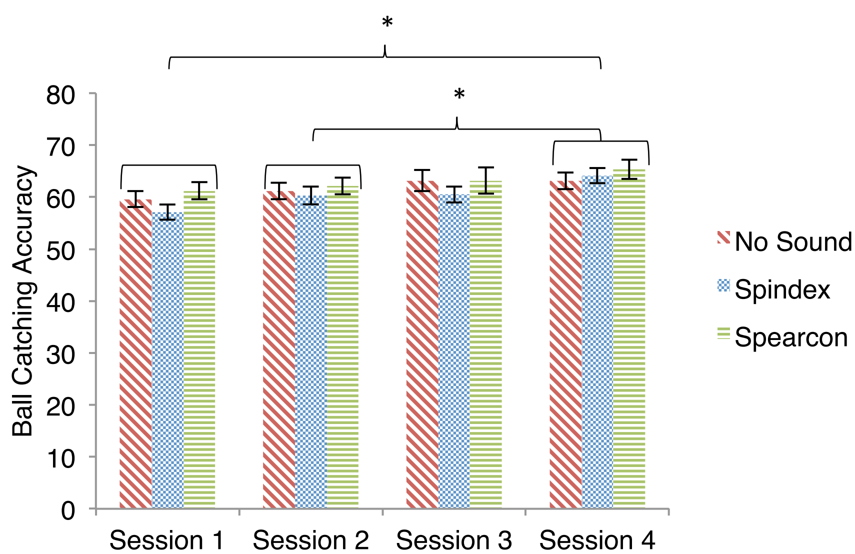


Figure 3.24 - Graph of the mean ball catching accuracy for each condition across the training block averages for the 4 sessions. Standard error is shown via error bars and significant differences are marked with "*".

Balls Released. For the number of balls released the analysis revealed a significant main effect of session. The follow-up tests revealed that participants had significantly higher average balls released per block for each session, as the experiment went on. There was no significant difference in regards to the main effect of condition, nor any significant interaction. These analyses can be seen in Table 3.28 and the data are visualized in Figure 3.25 below.

Table 3.28 - Analysis table for mean balls released for each condition across the mean of the training blocks for the four sessions. Significant differences are marked with "*".

Balls Released				
ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	95.96	2.62, 117.97	< .001*	.681
Main effect of Condition	< 0.01	2.00, 45.00	.999	< .001
Interaction	0.37	5.24, 117.97	.877	.016.
Post-hoc t-test Results	*t*	df	*p*	
Session 1 – Session 2	-9.12	47.00	< .001*	
Session 1 – Session 3	-10.94	47.00	< .001*	
Session 1 – Session 4	-13.06	47.00	< .001*	
Session 2 – Session 3	-5.29	47.00	< .001*	
Session 2 – Session 4	-8.64	47.00	< .001*	
Session 3 – Session 4	-4.56	47.00	< .001*	

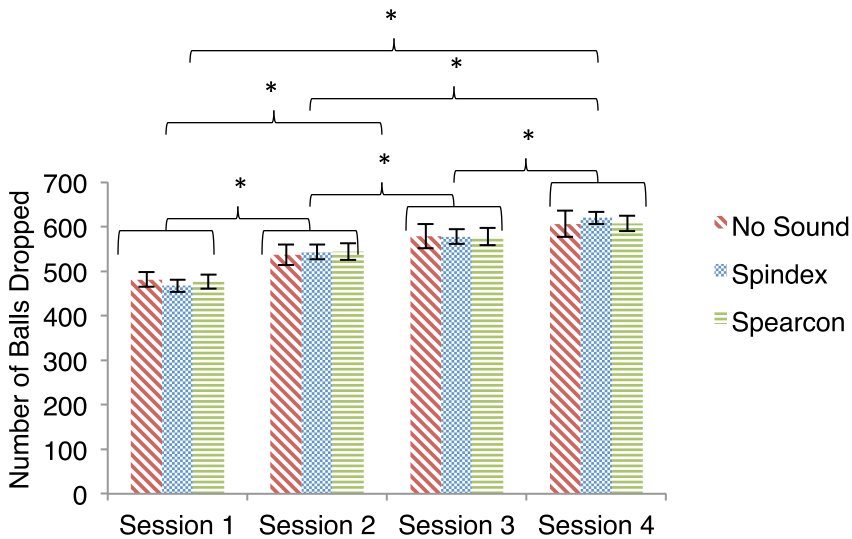


Figure 3.25 - Graph of the mean number of balls released for each condition across the training block averages for the 4 sessions. Standard error is shown via error bars and significant differences are marked with "*".

Training Blocks Ball Drop Performance Summary. The results seen in the data suggest that participants increased their abilities each session to catch more balls and in Session 4 increased their abilities to catch a higher percent of the balls released.

3.4.3 Training Block Visual Behaviors

For the 13 training blocks the raw eye-tracking output was used due to a lack of any baselines to compare against. The three measures used were the same measures used for the driving scenarios including time off rate (percent time off the ball drop task), glance count rate (average number of glances off the ball drop task per minute), and average dwell length (average length of time per dwell off the ball drop task).

Time Off Rate. For rate of time off the ball drop task there was a significant main effect of block, with the post-hoc revealing a that participants had a higher average time off rate per block in Session 1 than 2. There was no significant main effect of condition, nor a significant interaction. These analyses can be seen in Table 3.29 and the data are visualized in Figure 3.26 below.

Table 3.29 - Analysis table for mean time off rate for each condition across the mean of the training blocks for the four sessions. Significant differences are marked with "*".

Time off Rate				
ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	3.22	2.72, 119.71	.029*	.068
Main effect of Condition	0.01	2.00, 44.00	.992	< .001
Interaction	1.17	5.44, 119.71	.330	.050
Post-hoc t-test Results	*t*	df	*p*	
Session 1 – Session 2	2.77	46.00	.008*	
Session 1 – Session 3	1.74	46.00	.089	
Session 1 – Session 4	2.23	46.00	.031	
Session 2 – Session 3	-0.54	46.00	.594	
Session 2 – Session 4	0.20	46.00	.843	
Session 3 – Session 4	0.71	46.00	.482	

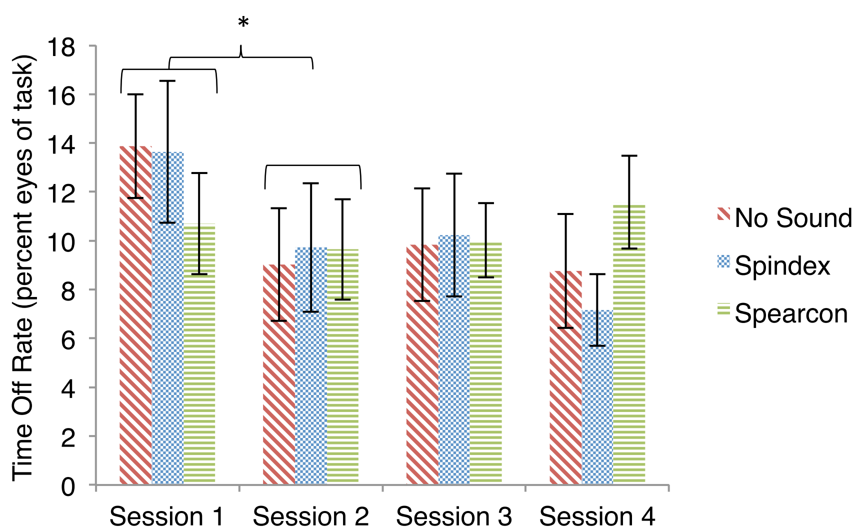


Figure 3.26 - Graph of the mean time off rate (percentage) for each condition across the training block averages for the 4 sessions. Standard error is shown via error

bars and significant differences are marked with "*".

Glance Count Rate. In regards to glance count rate there was a significant main effect of block. Post-hocs revealed that participants had a higher per block average glance count rate in Session 1 than in Sessions 2, 3, or 4. The analysis revealed no significant main effect of condition, nor a significant interaction. These analyses can be seen in Table 3.30 and the data are visualized in Figure 3.27 below.

Table 3.30 - Analysis table for mean glance count rate for each condition across the mean of the training blocks for the four sessions. Significant differences are marked with "*".

Glance Count Rate				
ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	12.19	2.94, 123.49	< .001*	.225
Main effect of Condition	0.39	2.00, 42.00	.677	.018
Interaction	0.42	5.88, 123.49	.863	.019
Post-hoc t-test Results	*t*	df	*p*	
Session 1 – Session 2	3.72	44.00	.001*	
Session 1 – Session 3	4.67	44.00	< .001*	
Session 1 – Session 4	4.83	44.00	< .001*	
Session 2 – Session 3	0.97	44.00	.335	
Session 2 – Session 4	1.22	44.00	.231	
Session 3 – Session 4	0.29	44.00	.775	

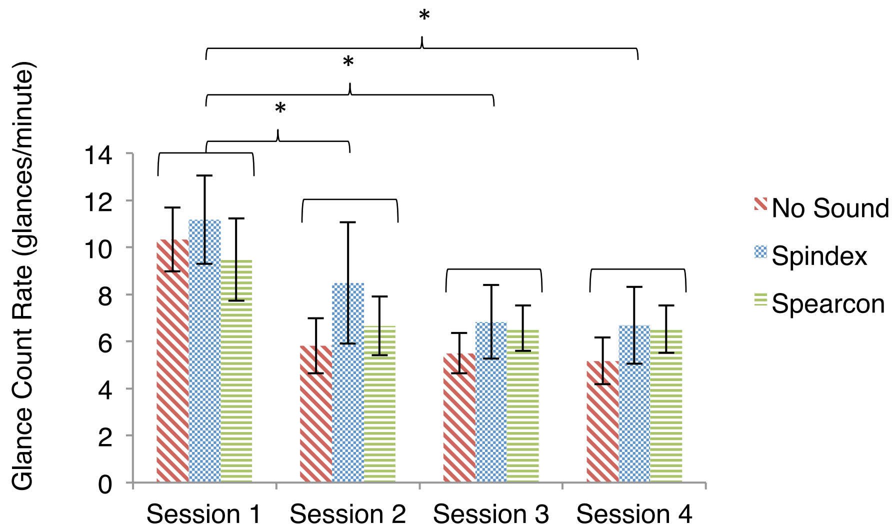


Figure 3.27 - Graph of the glance count rate (glances/minute) for each condition across the training block averages for the 4 sessions. Standard error is shown via error bars and significant differences are marked with "*".

Average Dwell Length. For average dwell length off the ball drop task there was a significant main effect of session with the analyses showing than participants had a higher average dwell length per block in Session 3 and 4 than in Session 1. The analysis revealed no significant main effect of condition, nor a significant interaction. These analyses can be seen in Table 3.31 and the data are visualized in Figure 3.28 below.

Table 3.31 - Analysis table for average dwell length for each condition across the mean of the training blocks for the four sessions. Significant differences are marked with "*".

Average Dwell Length

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	5.69	3.00, 135.00	.001*	.112
Main effect of Condition	1.68	2.00, 45.00	.199	.069
Interaction	1.08	6.00, 135.00	.380	.046
Post-hoc t-test Results	*t*	df	*p*	

Session 1 – Session 2	-2.51	47.00	.016
Session 1 – Session 3	-3.73	47.00	.001*
Session 1 – Session 4	-2.90	47.00	.006*
Session 2 – Session 3	-2.07	47.00	.044
Session 2 – Session 4	-0.94	47.00	.352
Session 3 – Session 4	0.71	47.00	.479

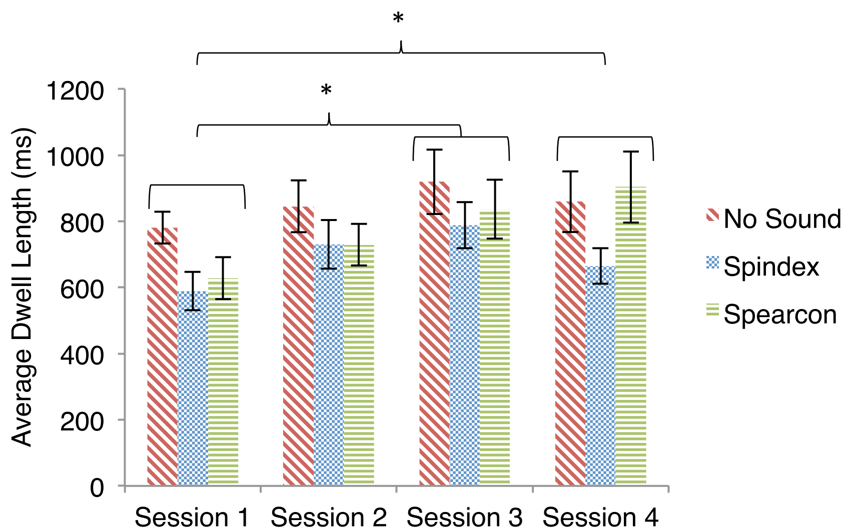


Figure 3.28 - Graph of the mean dwell length for each condition across the training block averages for the 4 sessions. Standard error is shown via error bars and significant differences are marked with "*".

Training Blocks Visual Behaviors Summary. The results seen here suggest less overall time off the primary task as time went on, but potentially longer glances off the primary task.

3.4.4 Training Block Objective Workload

Due to the previously seen differences in the heart rate values across the baseline measures, the same types of percent change differences scores were used in the analyses across training blocks for mean HR and HRV.

Percent Change in Mean Heart Rate. The analysis of percent change for mean HR revealed no significant main effect of block, condition, nor a significant interaction. These analyses can be seen in Table 3.32 and the data are visualized in Figure 3.29 below.

Table 3.32 - Analysis table for percent change in mean heart rate for each condition across the mean of the training blocks for the four sessions.

Percent Change in Mean Heart Rate

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	0.40	2.49, 104.72	.716	.009
Main effect of Condition	0.72	2.00, 42.00	.493	.033
Interaction	1.91	4.99, 104.72	.100	.083

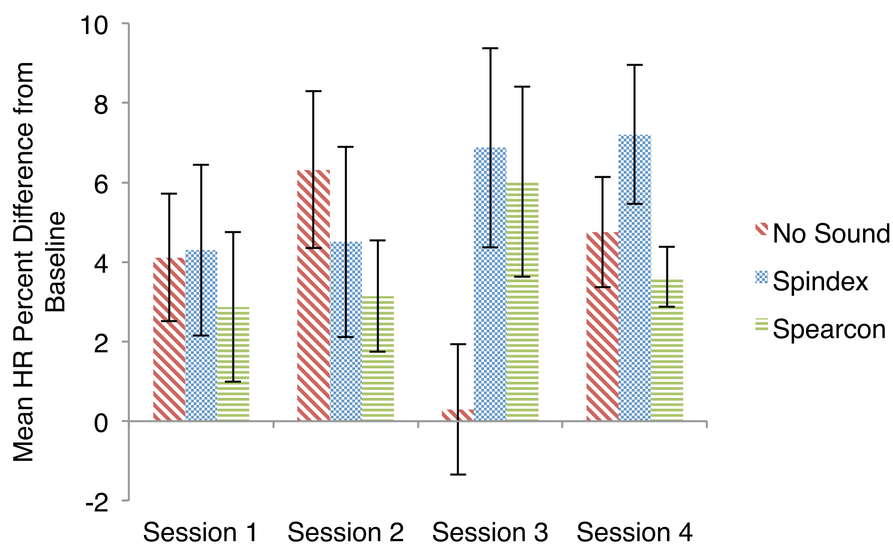


Figure 3.29 - Graph of the percent change in mean HR for each condition across the training block averages for the 4 sessions. Standard error is shown via error bars.

Percent Change in Mean Heart Rate Variability. The analysis of percent change for mean HRV also revealed no significant main effect of block, condition, nor a significant interaction. These analyses can be seen in Table 3.33 and the data are visualized in Figure 3.30 below.

Table 3.33 - Analysis table for percent change in mean heart rate variance for each condition across the mean of the training blocks for the four sessions.

Percent Change in Mean Heart Rate Variability

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	0.07	2.31, 87.60	.953	.002
Main effect of Condition	0.71	2.00, 38.00	.500	.036
Interaction	0.51	4.61, 87.60	.752	.026

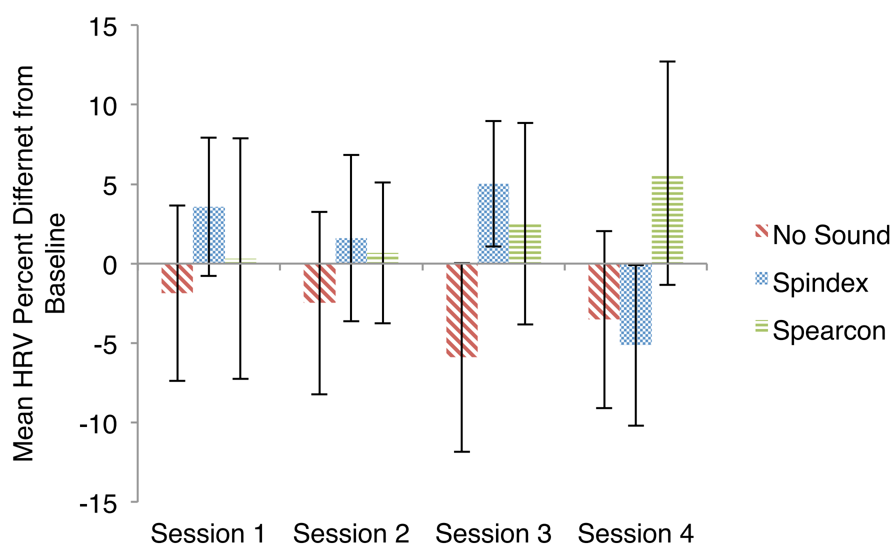


Figure 3.30 - Graph of the percent change in mean HRV for each condition across the training block averages for the 4 sessions. Standard error is shown via error bars.

Training Blocks Objective Workload Summary. The lack of any differences across the training blocks suggests that participants may have been working at peak performance across all blocks. This would make sense with the ball drop game being calibrated each session to increase the difficulty. The same lack of any differences between conditions as in the drives again suggests the lack of enough physiological response to the different conditions from the participants.

3.4.5 Training Block Subjective Workload

As with the subjective workload for the driving blocks, the analyses were done for each subscale and the composite score across the session averages.

Mental Workload. For mental workload a significant difference was found for the main effect of session, with the t-tests revealing participants had a higher block average mental workload in Session 1 than in Sessions 2, 3, or 4, and higher in Session 2 than in Session 4. No significant main effect of condition was found, nor was a significant interaction. These analyses can be seen in Table 3.34 and the data are visualized in Figure 3.31 below.

Table 3.34 - Analysis table for subjective mental workload NASA_TLX score for each condition across the mean of the training blocks for the four sessions. Significant differences are marked with "*".

Mental Workload

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	21.92	1.75, 76.94	< .001*	.332
Main effect of Condition	0.04	2.00, 44.00	.957	.002
Interaction	1.69	3.50, 76.94	.168	.071
Post-hoc t-test Results	*t*	df	*p*	
Session 1 – Session 2	50.00	46.00	< .001*	
Session 1 – Session 3	4.61	46.00	< .001*	
Session 1 – Session 4	5.27	46.00	< .001*	
Session 2 – Session 3	1.97	46.00	.055	
Session 2 – Session 4	3.31	46.00	.002*	

Session 3 – Session 4	2.37	46.00	.022

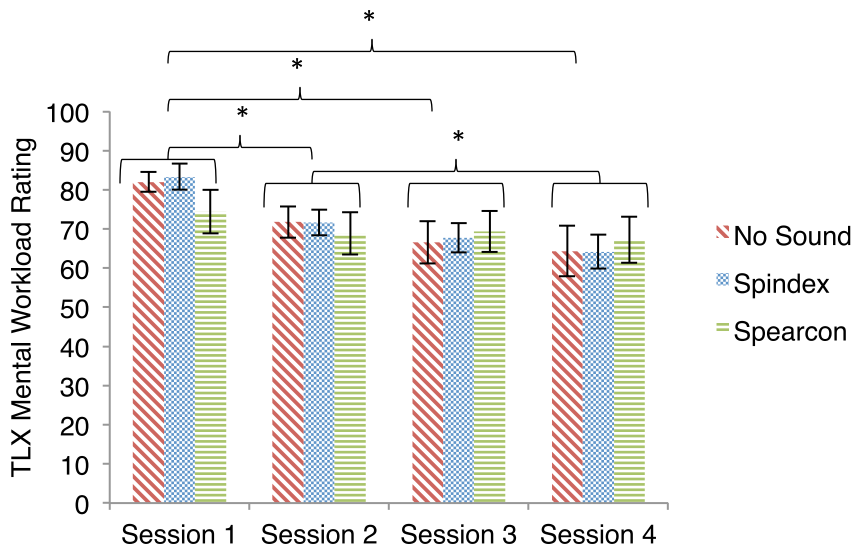


Figure 3.31 - Graph of the mean subjective NASA-TLX mental workload rating for each condition across the training block averages for the 4 sessions. Standard error is shown via error bars and significant differences are marked with "*".

Physical Workload. For physical workload the analysis revealed that participants ratings had no significant main effect of block, no significant main effect of condition, nor a significant interaction. These analyses can be seen in Table 3.35 and the data are visualized in Figure 3.32 below.

Table 3.35 - Analysis table for subjective physical workload NASA_TLX score for each condition across the mean of the training blocks for the four sessions.

Physical Workload

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	2.67	1.64, 80.58	.086	.052
Main effect of Condition	2.60	2.00, 49.00	.085	.096
Interaction	0.42	3.29, 80.58	.758	.017

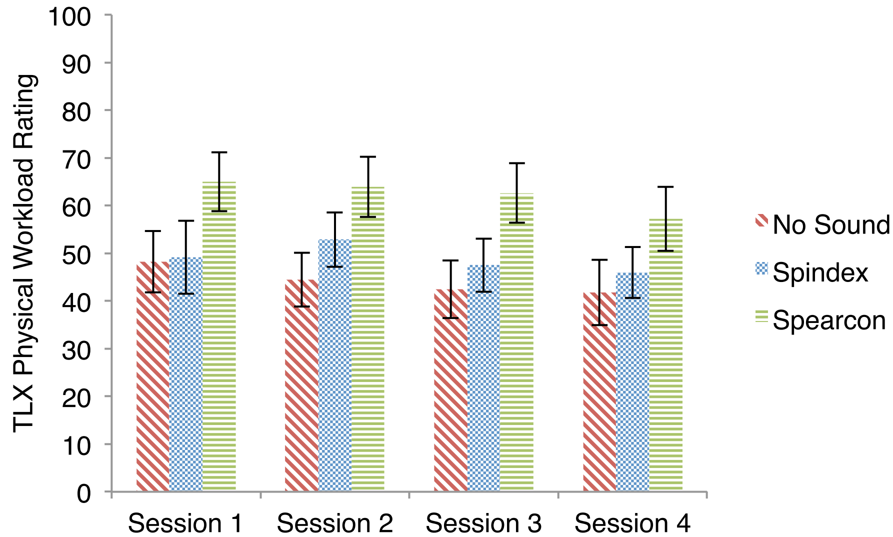


Figure 3.32 - Graph of the mean subjective NASA-TLX physical workload rating for each condition across the training block averages for the 4 sessions. Standard error is shown via error bars.

Subjective Performance Ratings. For subjective ratings of performance the analysis revealed a significant main effect of block, with the t-tests revealing that participants had a lower per block session average rating of their performance in Session 4 than 2. Although approaching significance there was no significant main effect of condition, nor a significant interaction. These analyses can be seen in Table 3.36 and the data are visualized in Figure 3.33 below.

Table 3.36 - Analysis table for subjective performance NASA_TLX score for each condition across the mean of the training blocks for the four sessions. Significant differences are marked with "*".

Performance

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	4.37	1.79, 84.33	.019*	.085
Main effect of Condition	2.90	2.00, 47.00	.063	.111

Interaction	1.35	3.59, 84.33	.262	.054
Post-hoc t-test Results	*t*	df	*p*	
Session 1 – Session 2	0.85	49.00	.402	
Session 1 – Session 3	1.62	49.00	.113	
Session 1 – Session 4	2.49	49.00	.016	
Session 2 – Session 3	1.84	49.00	.072	
Session 2 – Session 4	3.20	49.00	.002*	
Session 3 – Session 4	2.41	49.00	.020	

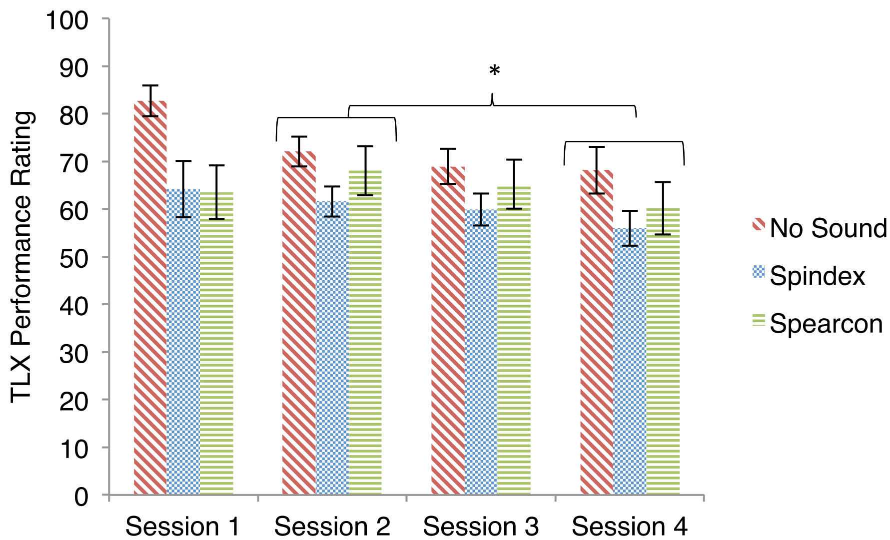


Figure 3.33 - Graph of the mean subjective NASA-TLX performance rating for each condition across the training block averages for the 4 sessions. Standard error is shown via error bars and significant differences are marked with "*".

Effort. In regards to effort participants' ratings showed a significant main effect of block. The post-hocs revealed that participants had a higher per block average rating of effort in Session 1 than Session 2, 3, or 4, and a higher rating in Session 2 than Session 3 or 4. The analysis found no significant main effect of condition, nor a significant interaction. These analyses can be seen in Table 3.37 and the data are visualized in Figure

3.34 below.

Table 3.37 - Analysis table for subjective effort NASA_TLX score for each condition across the mean of the training blocks for the four sessions. Significant differences are marked with "*".

Effort

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	33.61	1.96, 86.22	< .001*	.433
Main effect of Condition	1.26	2.00, 44.00	.295	.054
Interaction	1.82	3.92, 86.22	.133	.076
Post-hoc t-test Results	*t*	df	*p*	
Session 1 – Session 2	6.28	46.00	< .001*	
Session 1 – Session 3	6.64	46.00	< .001*	
Session 1 – Session 4	6.56	46.00	< .001*	
Session 2 – Session 3	2.82	46.00	.007*	
Session 2 – Session 4	2.88	46.00	.006*	
Session 3 – Session 4	0.87	46.00	.390	

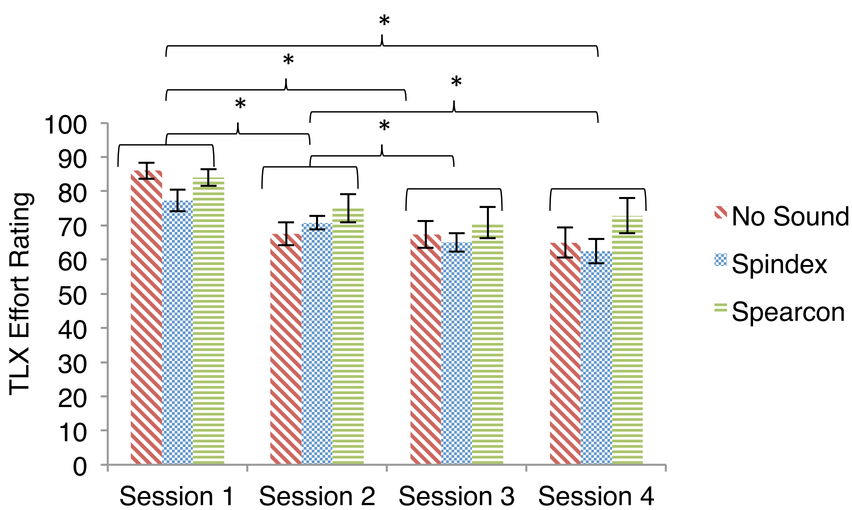


Figure 3.34 - Graph of the mean subjective NASA-TLX effort rating for each condition across the training block averages for the 4 sessions. Standard error is shown via error bars and significant differences are marked with "*".

Temporal Workload. For temporal workload participant ratings results in a significant main effect of block, with post-hocs revealing participants' higher per block average rating of temporal workload in Session 1 than Session 2, 3, or 4. The analysis revealed no significant main effect of condition, nor any significant interaction. These analyses can be seen in Table 3.38 and the data are visualized in Figure 3.35 below.

Table 3.38 - Analysis table for subjective temporal workload NASA_TLX score for each condition across the mean of the training blocks for the four sessions. Significant differences are marked with "*".

Temporal Workload				
ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	9.90	1.85, 85.13	< .001*	.177
Main effect of Condition	1.24	2.00, 46.00	.300	.051
Interaction	1.29	3.70, 85.13	.281	.053
Post-hoc t-test Results	*t*	df	*p*	
Session 1 – Session 2	2.81	48.00	.007*	
Session 1 – Session 3	3.65	48.00	.001*	
Session 1 – Session 4	3.40	48.00	.001*	
Session 2 – Session 3	2.00	48.00	.051	
Session 2 – Session 4	1.77	48.00	.084	
Session 3 – Session 4	0.35	48.00	.728	

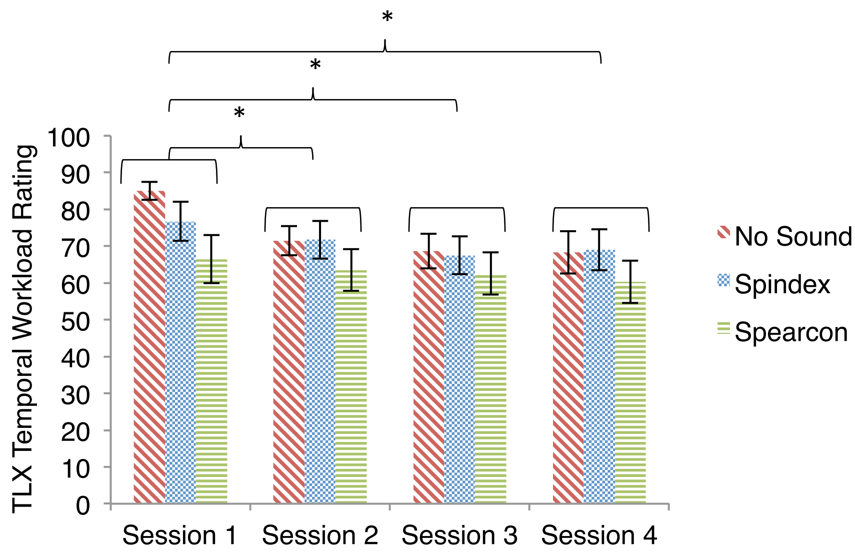


Figure 3.35 - Graph of the mean subjective NASA-TLX temporal workload rating for each condition across the training block averages for the 4 sessions. Standard error is shown via error bars and significant differences are marked with "*".

Frustration. Analysis of the frustration ratings in the training blocks revealed a significant main effect of block. Paired t-tests revealed that participants had a higher per block average rating of frustration in Session 1 than Sessions 2, 3, or 4, and higher rating in Session 2 than Sessions 3 or 4. There was no main effect of condition found in the analysis, however there was a significant interaction found. The significant interaction was investigated via one-way ANOVAs and interactions were found to be between the sessions in the No Sound condition and the sessions in the Spindex condition. The t-tests to investigate these differences found that for the No Sound condition there was a significant difference between Session 1 and Session 2, 3, and 4, with frustration being higher for Session 1 than the other sessions. For the Spindex condition the t-tests revealed a significant difference in average frustration ratings in the training blocks for Session 1 and Session 4, with Session 1 having higher scores. These analyses can be seen in Table

3.39 and the data are visualized in Figure 3.36 below.

Table 3.39 - Analysis table for subjective frustration NASA-TLX score for each condition across the mean of the training blocks for the four sessions. Significant differences are marked with "*".

Frustration				
ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	44.20	2.36, 108.47	< .001*	.490
Main effect of Condition	1.78	2.00, 46.00	.180	.072
Interaction	4.69	4.72, 108.47	.001*	.169
Post-hoc t-test Results	*t*	df	*p*	
Session 1 – Session 2	4.00	48.00	< .001*	
Session 1 – Session 3	6.48	48.00	< .001*	
Session 1 – Session 4	7.40	48.00	< .001*	
Session 2 – Session 3	5.05	48.00	< .001*	
Session 2 – Session 4	6.40	48.00	< .001*	
Session 3 – Session 4	2.40	48.00	.020	
Interaction Analyses	*F*	*df*	*p*	
Session 1 x Condition	1.91	2.00, 48.00	.160	
Session 2 x Condition	1.33	2.00, 49.00	.274	
Session 3 x Condition	2.11	2.00, 48.00	.133	
Session 4 x Condition	2.79	2.00, 47.00	.072	
No Sound x Session	9.02	3.00, 61.00	< .001*	
Spindex x Session	4.48	3.00, 68.00	.006*	
Spearcon x Session	0.55	3.00, 64.00	.650	
Interaction Post-hoc t-tests	*t*	df	*p*	
No Sound: Session 1-Session 2	2.94	31.00	.006*	
No Sound: Session 1-Session 3	4.25	31.00	< .001*	
No Sound: Session 1-Session 4	5.64	29.00	< .001*	
No Sound: Session 2-Session 3	1.12	32.00	.240	
No Sound: Session 2-Session 4	2.16	30.00	.039	
No Sound: Session 3-Session 4	0.92	30.00	.366	

Spindex: Session 1-Session 2	1.33	34.00	.192
Spindex: Session 1-Session 3	2.51	34.00	.017
Spindex: Session 1-Session 4	3.16	34.00	.003*
Spindex: Session 2-Session 3	1.37	34.00	.179
Spindex: Session 2-Session 4	2.16	34.00	.038
Spindex: Session 3-Session 4	0.84	34.00	.406

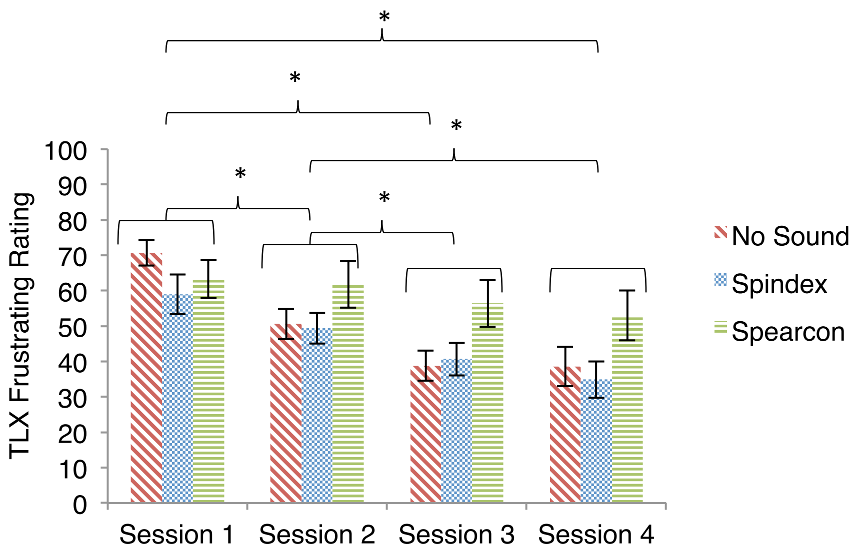


Figure 3.36 - Graph of the mean subjective NASA-TLX frustration rating for each condition across the training block averages for the 4 sessions. Standard error is shown via error bars and significant differences are marked with "*".

NASA-TLX Composite Workload. For the total NASA-TLX workload ratings the analysis found a significant main effect of block. This finding was followed up with t-tests, which revealed that participants had a higher per block average rating of composite workload in Session 1 than Sessions 2, 3, or 4, and higher rating in Session 2 than Sessions 3 or 4. The analysis revealed no significant differences between conditions, and no significant interaction between block and condition. These analyses can be seen in Table 3.40 and the data are visualized in Figure 3.37 below.

Table 3.40 - Analysis table for subjective composite workload NASA_TLX score for each condition across the mean of the training blocks for the four sessions. Significant differences are marked with "*".

Composite Workload

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	27.20	1.98, 94.97	< .001*	.362
Main effect of Condition	0.01	2.00, 48.00	.991	< .001
Interaction	1.32	3.96, 94.97	.270	.052

Post-hoc t-test Results	*t*	df	*p*
Session 1 – Session 2	5.08	50.00	< .001*
Session 1 – Session 3	5.66	50.00	< .001*
Session 1 – Session 4	6.16	50.00	< .001*
Session 2 – Session 3	2.74	50.00	.008*
Session 2 – Session 4	3.77	50.00	< .001*
Session 3 – Session 4	2.17	50.00	.035

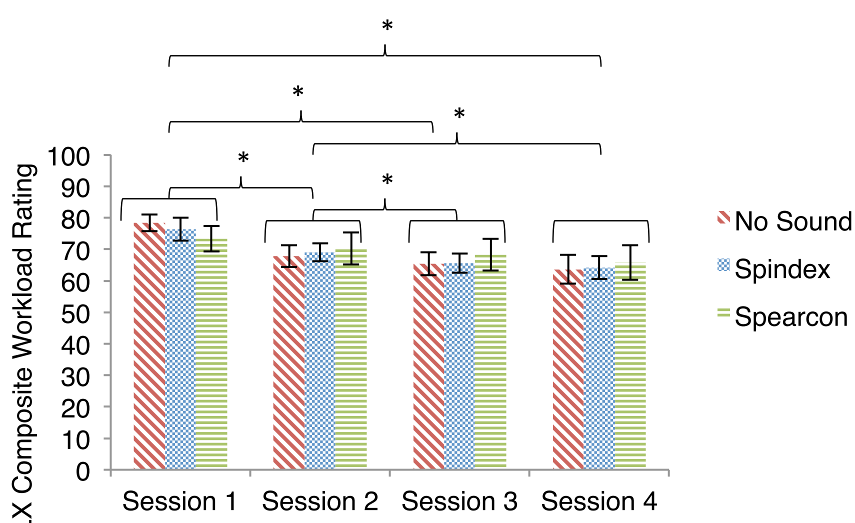


Figure 3.37 - Graph of the mean subjective NASA-TLX composite workload rating for each condition across the training block averages for the 4 sessions. Standard error is shown via error bars and significant differences are marked with "*".

Training Blocks Subjective Workload Summary. The results seen here suggest that the block average workload rating for each session decreased across the types of workload as participants went through the experiment. The perceived *performance* rating across blocks seemed to decrease, but this may have been due to a common mistake in filling out the perceived performance scale backwards. The interactions for frustration show a trend of decreasing frustration across the No Sound and Spindex conditions, but at different rates, with no decrease in frustration for the Spearcon condition over the period of the study.

3.4.6 Preferences and Perceived Performance

The preference measures from the questionnaires at the end of each session included questions of the participant's effectiveness at the search task, their effectiveness at the primary task(s), and whether the display was effective, helpful, or annoying. These measures were collected across all four sessions on a Likert-like scale from 1 to 6 and the descriptive data can be seen in Table J.19 in Appendix J. Participants were also asked to describe their approach to completing the dual task scenario and if it changed between Sessions 1 and 3. The analyses of these data included all the data points, as no outliers were removed from the preference data due to limited range of values from the Likert-like data. All main effect post-hocs for the factor of session were compared using a corrected alpha of .008 (.05/6 comparisons).

Effective at Search Task. For participants' ratings of their effectiveness at the search task the ANOVA revealed a significant main effect of session. This was followed up by t-tests, which revealed that participants reported significantly lower ratings of effectiveness for the search task on Session 1 than on Sessions 3 or Session 4, lower scores on Session 2 than 3 or Session 4, and lower scores on Session 3 than Session 4.

There was no significant main effect of condition found in the analysis, but there was a significant interaction.

The post-hoc ANOVAs for the interaction revealed potential significant interactions between the conditions in Session 4, between the sessions in the Spindex condition, and between the sessions in the Spearcon condition. For the conditions in Session 4 the t-tests revealed significant differences between the No Sound condition and Spindex condition and the No Sound and Spearcon condition, with the No Sound condition having a lower rating of effectiveness at the search task in that session. For the sessions in the Spindex condition the t-tests revealed that both Session 1 and Session 2 were significantly different from Session 3 and Session 4, with participants in the Spindex condition feeling they were better in Session 3 and Session 4 than the first two sessions. Finally, for the Spearcon condition participants rated their effectiveness as significantly greater in Session 4 than Session 1, 2, or 3. These analyses can be seen in Table 3.41 and the data are visualized in Figure 3.38 below.

Table 3.41 - Analysis table for perceived effectiveness at the search task for each condition across the four sessions. Significant differences are marked with "*".

Effective at Search Task

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	33.58	3.00, 147.00	< .001*	.407
Main effect of Condition	2.65	2.00, 49.00	.081	.098
Interaction	2.76	6.00, 147.00	.014*	.101
Post-hoc t-test Results	*t*	df	*p*	
Session 1 – Session 2	1.24	51.00	.220	
Session 1 – Session 3	-3.57	51.00	.001*	
Session 1 – Session 4	-7.00	51.00	< .001*	
Session 2 – Session 3	-4.93	51.00	< .001*	

Session 2 – Session 4	-8.52	51.00	< .001*
Session 3 – Session 4	-3.98	51.00	< .001*
Interaction Analyses	*F*	*df*	*p*
Session 1 x Condition	0.30	2.00, 49.00	.743
Session 2 x Condition	0.89	2.00, 49.00	.416
Session 3 x Condition	4.08	2.00, 49.00	.023
Session 4 x Condition	6.75	2.00, 49.00	.003*
No Sound x Session	1.87	3.00, 64.00	.143
Spindex x Session	10.44	3.00, 68.00	< .001*
Spearcon x Session	9.25	3.00, 64.00	< .001*
Interaction Post-hoc t-tests	*t*	df	*p*
Session 4: No Sound-Spindex	2.99	33.00	.005*
Session 4: No Sound-Spearcon	3.20	32.00	.003*
Session 4: Spearcon-Spindex	0.16	33.00	.875
Spindex: Session 1-Session 2	0.18	34.00	.862
Spindex: Session 1-Session 3	3.51	34.00	.001*
Spindex: Session 1-Session 4	4.56	34.00	< .001*
Spindex: Session 2-Session 3	3.33	34.00	.002*
Spindex: Session 2-Session 4	4.27	34.00	< .001*
Spindex: Session 3-Session 4	0.99	34.00	.329
Spearcon: Session 1-Session 2	1.12	32.00	.271
Spearcon: Session 1-Session 3	0.70	32.00	.488
Spearcon: Session 1-Session 4	4.19	32.00	< .001*
Spearcon: Session 2-Session 3	1.82	32.00	.078
Spearcon: Session 2-Session 4	6.14	32.00	< .001*
Spearcon: Session 3-Session 4	3.12	32.00	.004*

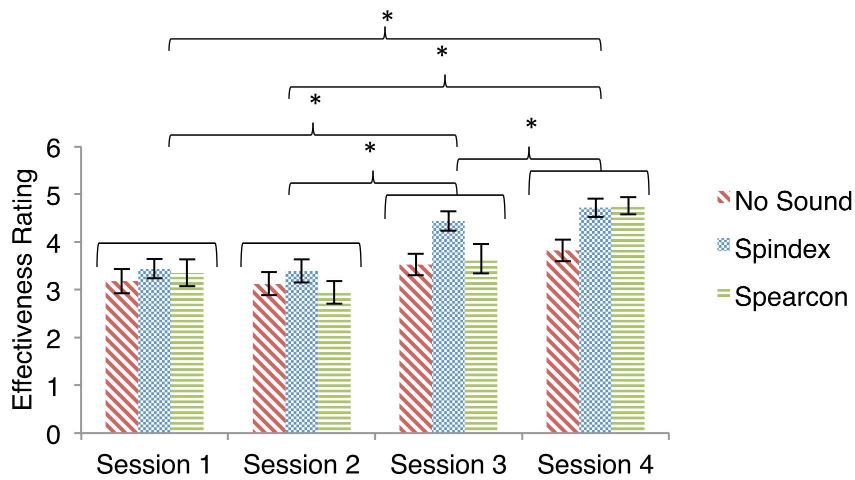


Figure 3.38 - Graph of the mean perceived self effectiveness at the search task for each condition across the four sessions. Standard error is shown via error bars and significant differences are marked with "*".

Display is Effective. For participants' ratings of the effectiveness of the display, the ANOVA revealed a significant main effect of session. Post-hocs revealed that participants rated the displays as significantly more effective on Session 4 than Session 1, Session 2, or Session 3. There was no significant main effect of condition, and no significant interaction. These analyses can be seen in Table 3.42 and the data are visualized in Figure 3.39 below.

Table 3.42 - Analysis table for perceived display effectiveness for each condition across the four sessions. Significant differences are marked with "*".

Display is Effective

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	8.51	3.00, 147.00	< .001*	.148
Main effect of Condition	0.68	2.00, 49.00	.509	.027
Interaction	1.64	6.00, 147.00	.140	.063
Post-hoc t-test Results	*t*	df	*p*	
Session 1 – Session 2	1.14	51.00	.261	

Session 1 – Session 3	-0.22	51.00	.828
Session 1 – Session 4	-3.37	51.00	.001*
Session 2 – Session 3	-1.38	51.00	.175
Session 2 – Session 4	-4.93	51.00	< .001*
Session 3 – Session 4	-4.18	51.00	< .001*

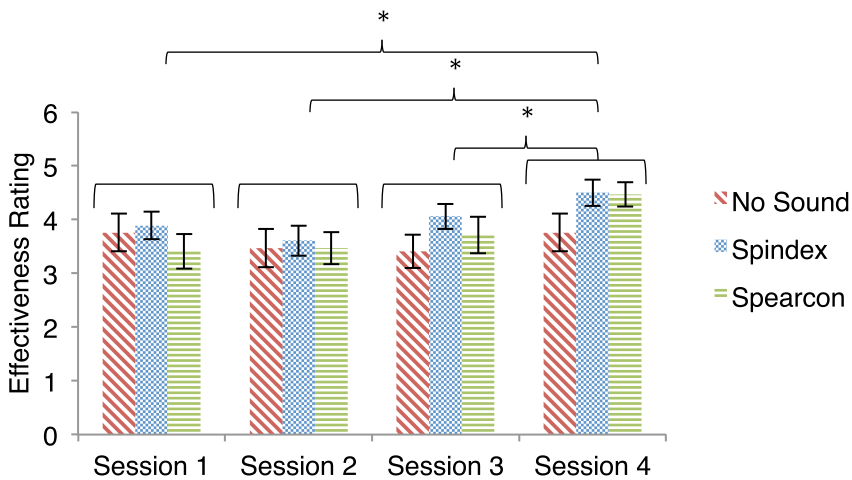


Figure 3.39 - Graph of the mean perceived display effectiveness for each condition across the four sessions. Standard error is shown via error bars and significant differences are marked with "*".

Display is Functionally Helpful. For participants ratings of whether the display was functionally helpful, there was a significant main effect of session, with the post-hocs showing that participants rated the displays as significantly more functionally helpful in Session 4 than in Session 1, Session 2, or Session 3. While no main effect of condition was found to be significant, the ANOVA did reveal a significant interaction, but the post-hoc ANOVAs revealed no significant interactions. These analyses can be seen in Table 3.43 and the data are visualized in Figure 3.40 below.

Table 3.43 - Analysis table for perceived functional helpfulness of the display for each condition across the four sessions. Significant differences are marked with "*".

Display is Functionally Helpful

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	8.06	2.82, 138.01	< .001*	.141
Main effect of Condition	1.71	2.00, 49.00	.192	.065
Interaction	2.80	5.64, 138.01	.015*	.103
Post-hoc t-test Results	*t*	df	*p*	
Session 1 – Session 2	0.38	51.00	.709	
Session 1 – Session 3	0.00	51.00	1.000	
Session 1 – Session 4	-3.08	51.00	.003*	
Session 2 – Session 3	-0.52	51.00	.606	
Session 2 – Session 4	-4.32	51.00	< .001*	
Session 3 – Session 4	-4.14	51.00	< .001*	
Interaction Analyses	*F*	*df*	*p*	
Session 1 x Condition	0.29	2.00, 49.00	.748	
Session 2 x Condition	3.50	2.00, 49.00	.038	
Session 3 x Condition	1.54	2.00, 49.00	.225	
Session 4 x Condition	2.54	2.00, 49.00	.089	
No Sound x Session	1.01	3.00, 64.00	.394	
Spindex x Session	1.33	3.00, 68.00	.273	
Spearcon x Session	0.60	3.00, 64.00	.603	

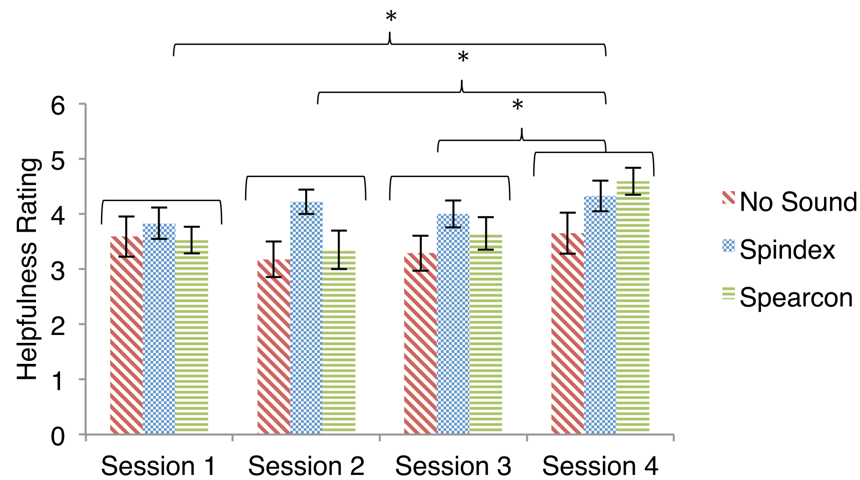


Figure 3.40 - Graph of the mean perceived functional helpfulness of the display for each condition across the four sessions. Standard error is shown via error bars and significant differences are marked with "*".

Display is Annoying. For whether the participants found the display as annoying there was a significant main effect of session, with the post-hocs showing that participants found the displays as significantly less annoying in Session 4 than Sessions 1, 2, or 3. While approaching the threshold, no main effect of condition was found to be significant, but there was a significant interaction, however none of the post-hoc ANOVAs was found to be significant. These analyses can be seen in Table 3.44 and the data are visualized in Figure 3.41 below.

Table 3.44 - Analysis table for perceived annoyance of the display for each condition across the four sessions. Significant differences are marked with "*".

Display is Annoying

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	8.01	2.85, 139.63	< .001*	.140
Main effect of Condition	2.77	2.00, 49.00	.072	.102
Interaction	2.97	5.70, 139.63	.011*	.108

Post-hoc t-test Results	*t*	df	*p*
Session 1 – Session 2	-0.22	51.00	.826
Session 1 – Session 3	0.90	51.00	.371
Session 1 – Session 4	4.30	51.00	< .001*
Session 2 – Session 3	1.02	51.00	.314
Session 2 – Session 4	3.37	51.00	.001*
Session 3 – Session 4	3.35	51.00	.002*
Interaction Analyses	***F***	***df***	***p***
Session 1 x Condition	2.65	2.00, 49.00	.081
Session 2 x Condition	4.82	2.00, 49.00	.012
Session 3 x Condition	4.02	2.00, 49.00	.024
Session 4 x Condition	0.02	2.00, 49.00	.978
No Sound x Session	2.10	3.00, 64.00	.110
Spindex x Session	0.67	3.00, 68.00	.574
Spearcon x Session	0.68	3.00, 64.00	.569

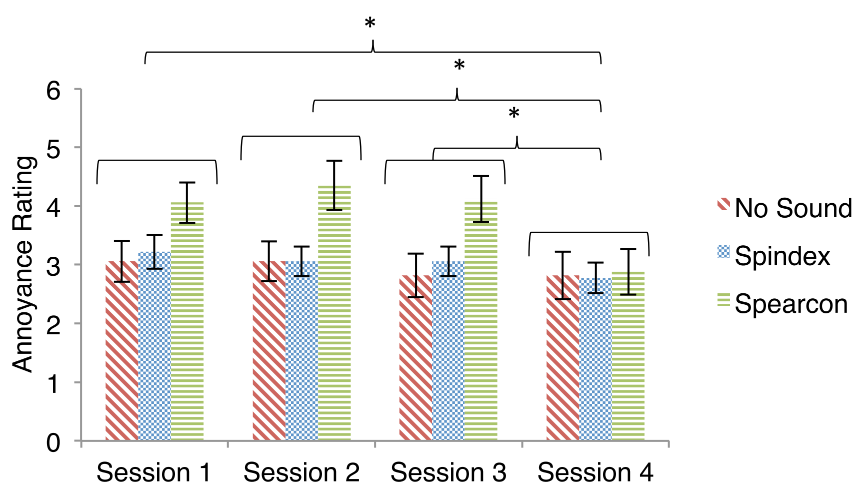


Figure 3.41- Graph of the mean perceived annoyance of the display for each condition across the four sessions. Standard error is shown via error bars and significant differences are marked with "*".

Effective At Primary Task. The ANOVA investigating participants' ratings for their effectiveness at the primary task revealed a significant main effect of session, with the post-hocs showing that participants rated themselves as being more effective in Session 4 than 1, 2, or 3. The ANOVA revealed no significant main effect of condition, nor a significant interaction. These analyses can be seen in Table 3.45 and the data are visualized in Figure 3.42 below.

Table 3.45 - Analysis table for perceived effectiveness at the primary task for each condition across the four sessions. Significant differences are marked with "*".

Effective at Primary Task

ANOVA Results	*F*	df	*p*	*η2*
Main effect of Session	29.19	2.89, 141.66	< .001*	.373
Main effect of Condition	2.05	2.00, 49.00	.139	.077
Interaction	1.09	5.78, 141.66	.373	.042
Post-hoc t-test Results	*t*	df	*p*	
Session 1 – Session 2	0.58	51.00	.563	
Session 1 – Session 3	0.11	51.00	.909	
Session 1 – Session 4	-8.35	51.00	< .001*	
Session 2 – Session 3	-0.56	51.00	.577	
Session 2 – Session 4	-8.94	51.00	< .001*	
Session 3 – Session 4	-6.88	51.00	< .001*	

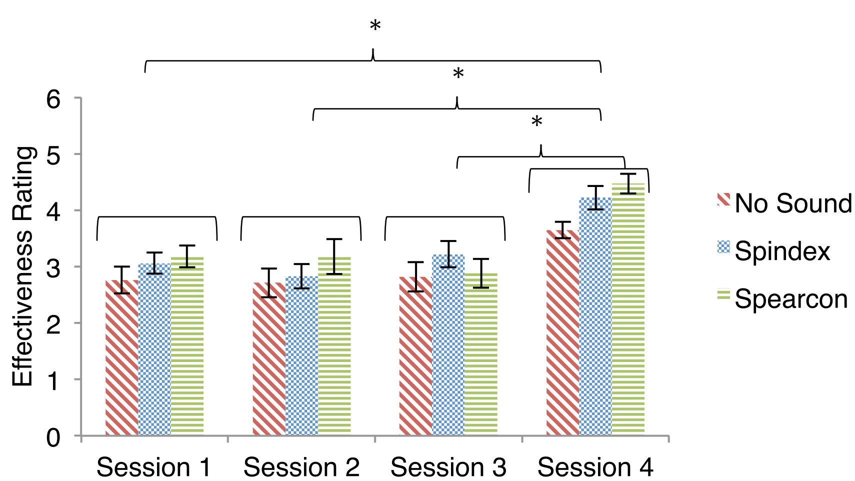


Figure 3.42- Graph of the mean perceived self-effectiveness at the primary task for

each condition across the four sessions. Standard error is shown via error bars and significant differences are marked with "*".

Change in Approach. When asked about the strategies participants were using at the end of the third session, participants had mixed answers. In the Spindex and Spearcon conditions, 9 and 13 participants, respectively, mentioned that they were relying more on the audio than they did at the beginning of the study. In these conditions only 3 Spindex participants and 2 Spearcon participants did not mention use of the audio at all. Other common answers included mixes of changes in scrolling strategies to try and make larger jumps towards the correct region the song was in, and the use of more peripheral vision for the tasks. Overall though all but 2 participants across the 3 conditions reported changing their approach in completing the tasks over time.

Preferences and Perceived Performance Summary. The results seen from these analyses show participants had higher preference and perceived performance with the interfaces on the last session as compared to all other sessions. There were no significant differences in regards to condition across all of the blocks but the significant interactions show that participants' opinions in some conditions changed at different rates than in others. The open answer regarding strategy changes over the study period revealed that a majority of the participants in the audio conditions used the audio more at the end of the study than at the beginning. However, it seemed that more of those in the Spearcon condition changed their strategy in regards to the audio than those in the Spindex condition. This could hint at more learning taking place in the Spearcon condition than in the Spindex condition, which may have been more usable from the beginning.

CHAPTER 4: DISCUSSION

4.1 Research Questions

The following sections aim to break down the results of the present study in regards to the research questions and hypotheses put forth. The focus here is to simply answer the research questions and hypotheses. The implications of these findings on the use of multimodal displays and the theory surrounding their use are discussed further in the sections after.

4.1.1 RQ2.1 - Practice Effects

The first research question was aimed at checking to see if there was a practice effect across the dependent variables as was expected from the pilot. Hypotheses 2.1.1-2.1.5 were tested through investigating the simple main effects of driving block and training session across the dependent measures. It should be noted that the increasing difficulty of the training task was a potential confound of the data for the training sessions and should be considered when viewing the results for any of training blocks. Similarly, the drives under main consideration here are those of Drive 1 and Drive 2, not necessarily Drive 3, as it was a more difficult drive and cannot be directly compared to the other drives.

H2.1.1 - Secondary Task Performance. H2.1.1 was that increased secondary task performance would be seen over time, through increased accuracy and decreased search times. For the driving blocks this was not supported, as while there was a significant main effect of drive on both selection time and number of searches across Drives 1 and 2, it was in the opposite direction as hypothesized. This was potentially due to a focus of attention on the driving task instead of the search task in the last session as there was no change in the percent correct selections measure. This suggests that there may not have been a decrease in abilities, but a shift in priorities, also called willingness to engage (Ranney, Mazzae, Garrott, & Goodman, 2000). This potential change in willingness to engage is further supported by a similarly significant difference between Drive 3 and Drives 1 and 2, with participants searching fewer times and having a higher average time to find a song in Drive 3 but again no decrease in percentage accuracy. These differences could suggest that participants simply found the driving task to be more important to perform safely in Drive 2 than Drive 1.

For the training blocks however there did seem to be a learning effect present, even with the increasingly more difficult ball drop task. Significant main effects were seen for the number of searches performed, with participants having a lower number of average searches per block in Session 1 than another other session, and then having fewer searches in Session 2 than in Session 4. In regards to percent correct, participants had higher average accuracy per block in Sessions 2, 3, and 4 than Session 1. Time to find a song also suggested an effect of practice, with participants having a lower average search time per block in Sessions 3 and 4 than Session 1, and in Session 4 than Session 2.

These results suggest that a practice effect for the search task took place in the training blocks but not the driving blocks. This leaves the question as to why these differences were not seen in the driving blocks. It could simply be that participants did not get enough practice with the driving task and that the skills gained in the ball drop task are not directly applicable to a driving task. However, then one would expect to simply see no change in the performance across Drives 1 and 2. Instead it seems that there was a change in willingness to engage and participants focused more of their attention on the driving task instead of the cell phone task as they perceived it to be of higher importance or more difficult than the search task.

H2.1.2 – Primary Task Performance. Although a potential confound, primary task performance was also of interest for learning effects. The hypothesis stated that the ball drop task performance measure of percent caught would see no change across the block averages for each session but that the number released would increase. In addition it was hypothesized that driving performance would improve in Drives 2 and compared to Drive 1.

For the driving performance, no differences were seen between Drives 1 and 2 for any of the measures. However there were differences between the first two drives and the third, more difficult drive, with better driving performance taking place in the first two drives. This points to the expected increased difficulty of Drive 3. For the training blocks, the number of balls released significantly increased for each session. This means that there were significantly more balls released in Session 4 than in Sessions 1, 2, or 3, more in Session 3 than Sessions 1, and 2, and more in Session 2 than Session 1. Surprisingly the percent accuracy also increased, with Session 4 having higher percent accuracy than Sessions 1 or 2.

These results show that practice did in fact increase the primary task performance for the ball-drop game, but not for the driving task. Whether the absence of differences for the driving task was due to lack of practice on the task itself or simply not having more resources to apply to that task is not clear. For the ball drop game the increased number of balls released shows that participants were getting better at the primary task over time. The percent accuracy also shows that they were getting better at the game in the dual-task situation and that a practice effect was taking place. The search task results show that this increase was not occurring at the cost of the search task and that performance was indeed improving across the dual task as a whole in the training blocks.

H2.1.3 - Visual Behaviors. For visual behaviors the learning effect was investigated again through the main effect of drives (mostly Drive 1 and 2 of interest) and training block averages across sessions. The hypothesis was that as practice increased participants would have less visual time off the road across the conditions. For the driving blocks the measures of percent time eyes off the road and count rate were both significantly lower in Drive 2 than Drive 1, pointing to less visual attention paid to the off-road task. This however does not necessarily mean better task sharing over time, as the participants also performed a lower number of searches during the second drive. For the training block session averages the time eyes off the primary task was significantly higher for Session 1 than Session 2 and the glance count rate was lower in Sessions 2, 3, and 4 than Session 1. Interestingly the dwell time off was higher is Session 4 and 3 than in Session 1, suggesting participants may have been choosing to have less glances away from the primary task but for longer periods of time. This was supported by a number of participants' qualitative feedback regarding any changes in strategy, with them stating they did less switching and instead simply got to the song required as quickly as possible with visual attention and then went back to the game.

The data associated with this hypothesis points to a general practice effect for the training task, with more practice seeming to decrease visual attention towards the secondary task. However the changes in difficulty for the training blocks for the ball drop may have confounded some of the data seen here and the practice effect may have been higher had this increase not taken place. The difference in the driving blocks is hard to explain due to the decreased number of searches done in Drive 2 as opposed to Drive 1 but suggests participants were able to better use the time they had with their eyes off the road after practicing on the training task. However, with previous work suggesting that even a small amount of time that drivers have their eyes off of the road being harmful, any difference could be meaningful (Klauer, Dingus, Neale, Sudweeks, & Ramsey, 2006). This points to the importance of eyes being on the road and the effect that this extra time with eyes on the primary task can have in a real world situation.

H2.1.4 - Workload. To investigate the practice effects through workload H2.1.2 hypothesized a decrease in subjective and objective workload across the sessions and drives. It should again be noted that the increasing difficulty of the primary training task could confound the data.

For the driving task a significant difference was found between Drives 1 and 2 for the objective workload measure of heart rate, with the results suggesting lower workload in Drive 2 than 1. In addition, subjective measures of mental workload, effort, temporal workload, frustration, and composite workload were lower in Drive 2 than Drive 1, suggesting decreases in workload across the two drives. There was however, also a decrease in subjective performance in Drive 2 than Drive 1. This could be explained by the actual decreased performance seen in the search task between the two drives but may also be contributed to a common mistake made by participants in filling out the performance question incorrectly in regards to performance due to its’ scale flipping as compared to the other measures, which was supported across the results seen elsewhere in the current study.

In regards to the training blocks, similar trends were seen as the driving blocks. While no objective workload differences were found, mental workload ratings were seen to be significantly lower in Sessions 2, 3, and 4 compared to Session 1, and lower in Session 4 than Session 2. For temporal workload Sessions 2, 3, and 4 were seen to have significantly lower average ratings than Session 1. Finally for effort, frustration, and composite workload participants responded with lower ratings in Session 2, 3, and 4 than Session 1, and lower ratings in Session 3 and 4 than Session 2. As seen in the driving blocks the performance ratings were against what would be expected, with a significant difference pointing to lower performance in Session 4 than Session 2.

These results point to a significant practice effect and support the training block results seen in secondary task performance. The results also support to the idea that the decreased secondary task performance in the drives was due to a decreased willingness to engage since the lower workload in the second drive as compared to Drive 1 implies that the decreased performance on the search task was not due to higher workload.

H2.1.5 - Perceived Performance. The final factor to investigate for a practice effect was that of perceived performance and preferences at the end of each session. For perceived effectiveness the participants rated themselves as significantly better in Sessions 3 and 4 than in Sessions 1 or 2, and better in Session 4 than Session 3. For the display being effective and for it being functionally helpful participants rated in the display higher in Session 4 than 1, 2, or 3. Similarly participants rated the annoyance of the display as significantly lower in Session 4 than Sessions 1, 2, or 3 and rated themselves as more effective at the primary task in Session 4 than 1, 2, or 3. These results support the idea that there was a learning effect with the displays and also clarifies that the performance TLX data seen in H2.1.2 was likely due to participant error in completing the survey.

RQ 2.1 Summary of Results. The results of RQ 2.1 point fairly clearly to a practice effect taking place. Almost all of the measures collected in the training blocks and many from the driving blocks point to learning effects occurring over time. However, a lack of differences in the performance on the primary driving task and the surprising decrease of performance on the secondary task is interesting. These results for the driving blocks can potentially be attributed to willingness to engage in the secondary task during these driving blocks. Even more interesting though, was that the participants improved their search task and primary task performance in the training task, even though the primary task increased in difficulty over time. This increase in performance over time suggests that there was some sort of process of automaticity taking place as described by Speelman and Kirsner (1997) where rule-based models evolved out of declarative steps, or as described by Schneider (1985) as a change in strategy. This is further supported when looking at the qualitative feedback regarding changing approaches for the participants, where all but 2 participants reported changing their approach over time. This result suggests that there were changes in strategies during the course of the study. Overall, these results suggest that there was a learning effect over the course of the study and that the hypotheses held that participants would improve their performance over time.

4.1.2 RQ2.2 – Effects of Condition in Training

The second research question investigated any differences between the conditions across the training sessions to determine if Hypotheses 2.2.1-2.2.5 held true. The hypotheses revolved around the law of practice (Newell & Rosenbloom, 1981), that as practice increased, with the participants with the multimodal displays were expected to increase their performance more so than those with the visual only display, due to their already present practice with visual displays. These hypotheses were tested by looking for any significant main effects of condition across the block averages by session or any interactions of Session x Condition for the dependent measures.

H2.2.1 – Secondary Task Performance. Hypothesis 2.2.1 was that secondary task performance would increase more so for those in the multimodal conditions. This was expected due to their potential for having more to gain from practice than the visuals-only condition participants, who were already potentially practiced at a search task using visuals. Support for this would be seen through interactions showing initially better or equal performance with the No Sound condition but then the auditory condition participants catching up or potentially surpassing the visual only participants in performance. However, for the secondary task measures there were no significant main effects of condition or interactions across the training sessions averages. This suggests no difference between the conditions were present initially, or as time went on, even as participants improved their performance across the three conditions as seen in the H2.1.1. This points to the participants having the same trend of learning across the 4 sessions.

H2.2.2 – Primary Task Performance. The performance on the primary task of ball drop was also a measure to investigate for differences between conditions. Looking for any differences in the calibration process or for the percentage caught during the trials could hint at differences between the conditions. There were however, no significant differences between the conditions or any interactions, pointing to no differences in the learning effects between the three conditions in this study.

H2.2.3 – Visual Behaviors. The hypothesis regarding visual behaviors suggested that the measures would decrease more so with practice for those in the auditory conditions, which would be seen by larger decreases in average percent eyes off task time, numbers of glances, and number of long glances off the primary task than those in the visuals only condition. However, no differences between conditions or interactions were found in the training block averages per session. This lack of differences may be due to the fact that participants did not yet feel conformable using the auditory displays, or that they already had high abilities to employ the auditory displays. The second possibility might suggest that practice with the auditory displays does not improve the use of said displays, which could point to the idea that MRT may hold without the need to consider practice.

H2.2.4 – Workload. For H2.2.4 similar expectations were in place as those for H2.2.1 but through decreased workload levels for the multimodal conditions across the training session averages. Support for the hypothesis would point to participants indeed learning more in the multimodal conditions than the visuals-only conditions as seen through lower workload levels.

Results from the analyses pointed to no differences in main effect or interactions for the objective workload measures of HR or HRV. While this goes against the hypothesis, this could be attributed to the confound of an increasingly more difficult primary task. For subjective workload there were no differences across the conditions or any interactions except for frustration. Significant interactions for frustration showed that participants in the No Sound condition reported having higher frustration in Session 1 than Session 2, 3, or 4. For the multimodal conditions the Spindex participants rated the display as less frustrating in Session 4 than in Session 1. These results suggest that participants become less frustrated with the No Sound display quickly, while they were less frustrated with the Spindex display only on the last day. Meanwhile no differences were seen in annoyance for those in the Spearcon condition, suggesting no change in the level of frustration with the display over time.

These results suggest that there were no real differences between conditions in subjective or objective workload outside of annoyance. It should be noted however that this, as with many of these measures, can be considered confounded due to the continually calibrated primary task difficulty. In order to investigate this, H.2.2.5 was used as a way to check if that was potentially a factor due to the expectation of the perceived performance measures being less likely to be affected by the calibration since participants could consider that as part of their self-ratings.

H2.2.5 – Perceived Performance. H2.2.5 was that there would be an interaction across blocks and conditions for perceived performance due to the additional learning that would occur for the multimodal conditions over time. This hypothesis aimed to serve as a check on the measures of workload and the effect that calibration of the primary task may have had on the overall ratings over time.

In investigating the perceived performance there was a significant interaction for participants' ratings of being effective at the search task. The post-hoc analyses point to there being a difference in Session 4, with the No Sound participants giving their display a significantly lower rating than participants in the Spindex or Spearcon conditions. This difference points to those in the two multimodal conditions feeling their displays were more effective in the final session than those in the No Sound condition. Furthermore, for the participants in the Spindex condition the ratings of display effectiveness were higher in Sessions 3 and 4 than in Session 1 and 2, suggesting that the participants thought of the display as more helpful in Sessions 3 and 4 than 1 or 2. Meanwhile, those in the Spearcon condition rated the display as more helpful in Session 4 than Sessions 1 or 2. This again points to increased perceived helpfulness of the multimodal displays, but only after Session 4 for the Spearcon condition. It should be noted that significant interactions were also found for perceived functional helpfulness and perceived annoyance but the post-hoc analyses found no significant differences.

The results found here of significantly higher perceived helpfulness of the display for the two multimodal conditions in the last session point to the multimodal condition participants believing the display was more helpful than those in the No Sound condition after the 4th session. This suggests that it either took participants 4 sessions to begin to perceived higher benefits of the multimodal displays or that something in the 4th session caused them to see an effectiveness of the multimodal display that they had not before. The interactions seen for the same measure for the Spindex and Spearcon conditions point towards the former, as the Spindex and Spearcon participants seemed to have a different rate of changing opinion, with those in the Spindex condition rating their display as more helpful than their initial reaction after Session 3. This is opposed to those in the Spearcon condition not reporting a difference until the 4th session. These results suggest that the rate of learning how to use the displays effectively, increased for those in the multimodal conditions as opposed to the No Sound condition and that it occurred at a different rate for the two multimodal conditions.

RQ2.2 Summary of Results. The results of the dependent measures across the training session averages point to only a few significant differences. The differences seen in the measures for the training sessions were the levels of annoyance and the perceived helpfulness of the display, which pointed towards No Sound and Spindex participants being less annoyed over time with their displays, and with participants rating the Spindex and Spearcon displays as more helpful as time went on. These results suggest no difference in the rate of practice effects took place for the multimodal conditions as compared to the visuals-only condition as hypothesized. This, along with the equal initial measures in the training blocks could suggest that more learning is needed for MRT to hold as is hypothesized in the current research or that the multimodal conditions are actually no better than the No Sound condition for this task.

However, what could be considered here is that the attempt at using the increasing difficulty of the primary ball drop task to control for learning effects on the ball drop game may have backfired and kept participants from learning how to effectively employ the auditory cues for the search task due to a focus on the ball drop task instead of the search task. If the difficulty of the ball drop task had not been increased each session, this may have allowed for participants to explore ways of interacting with the displays and using them to be more effective. However, differences for learning were seen over time but simply not across conditions, which show that practice effects were occurring even with the increasing difficulty of the primary training task. If this had not occurred then it could have been considered that constantly increasing difficulty of the primary task forced the participants to use their initial approach to completing the secondary task and focused more so on getting better at the primary task, therefore limiting the improvements that they made on the search task. However, the improvements made over time, across conditions, and the changes in approach discussed in the qualitative feedback, suggest that this was not the case, and instead that no major differences in learning rates occurred across the conditions in the training blocks. This would suggest that either MRT does not hold, and no differences existed between the participants in the different conditions, or that the participants were still learning how to improve on all of the displays during the training blocks.

4.1.3 RQ2.3 – Effects of Condition in Driving

The final research question, RQ2.3 was aimed at investigating the differences between conditions in the testing sessions after participants had reached a the determined amount of practice. Hypotheses 2.3.1-2.3.4 were tested through the 2x3 mixed ANOVAS across the dependent measures, focusing on the main effect of condition but considering the simple main effect of block and/or the interactions when necessary to explore further.

H2.3.1 – Secondary Task Performance. Regarding secondary search task performance, H2.3.1 suggested that accuracy and time to find a song would be higher for those users with the auditory cues than the visual-only cues. There were however, no differences between conditions for the number of trials, time to find a song, or in percent accuracy. As discussed earlier, the decrease across the three drives for the number of songs searched and average time to find a song suggested a change in willingness to engage in the driving task since percent accuracy did not change, but no differences across conditions were seen.

H2.3.2 – Driving Performance. For the driving task there were no differences between the conditions, which goes against the hypothesis of improved driving performance for those in the auditory conditions after training. However, there was a difference between the 3 drives, with participant's driving performance in Drive 3 having a higher mean distance gap, SD distance gap, and mean lane deviation, than in Drive 1 or 2, and higher SD lane deviation than Drive 2. These results suggest that Drive 3 was more difficult than Drives 1 or 2 as was expected, and suggests that participants' performance was different according to difficulty of the drive.

H2.3.3 – Visual Behaviors. The visual behavior measures discussed in H2.2.3 suggested that the users of the auditory displays would have less eyes off the road time, and less glances and long glances off the road as the visual-only participants as it was expected they would rely on the auditory cues to complete the tasks and less so on their visuals. The results of the analyses showed that for time off rate there was a significant difference between conditions, with those in the Spindex condition having a lower time off rate across drives than participants in either the No Sound or Spearcon conditions. For glance count rate there was also a significant difference, with those in the Spindex condition having lower rates than those in the No Sound condition. There was however no difference in average dwell length, suggesting participants did not change their dwell length differently across the conditions. These results support the reasoning of using these auditory cues given theoretically by authors such as Nees and Walker (2011) as people have more visual attention towards the driving task than if they used visual only displays.

H2.3.4 – Workload. For H2.3.4 the workload measures were hypothesized to be lower for those in the auditory conditions than the visuals-only condition as seen through lower heart rate and NASA-TLX scores. For objective workload there were no significant differences seen between conditions for either HR or HRV. There was however, significant a difference found for the subjective NASA-TLX metric of physical workload, with participants using the Spearcon condition having higher average ratings then those in the Spindex conditions. In addition, performance ratings were lower for those in the Spindex and Spearcon conditions than those in the No Sound condition, and lower in Spearcon than the Spindex conditions. However, it should be noted that as seen in some of the previous sections the performance data seemed to be confounded.

Outside of the physical workload differences seen between Spearcon and Spindex participant ratings, these results do not support the hypothesis and the results seen in previous work. In theory this suggests that MRT may not hold, or it could suggest that either not enough practice took place to decrease the workload differences or that it was too difficult of a dual task situation for differences to appear due to a floor effect.

RQ2.3 Summary of Results. The results of the hypotheses for this section point to a lack of differences between the conditions other than visual attention being more so on the road for the participants with the Spindex cues than the No Sound or Spearcon cues, and higher physical workload for those in the Spearcon condition as compared to those in the Spindex condition. These results support previous work in regards to visual behaviors (Gable, Walker, Moses, & Chitloor, 2013), with those using the Spindex cues improving their visual attention to the road as compared to those in the No Sound or Spearcons conditions. It should be noted here that while these differences for visual behaviors had no interactions, the changes in percent time off were much greater numerically for the multimodal cues than the No Sound condition. However, a lack of these differences being significant nor any other differences between conditions, particularly over time, points to a potential lack of significantly different practice effects as was expected. This lack of different practice effects across the conditions suggests that multimodal interfaces may not actually need to be highly practiced as was hypothesized and that MRT may hold without users becoming an expert at an interface type.

4.2 Implications of Results

4.2.1 Supporting Previous Work

The results of the study and the outcomes of the hypotheses are interesting when viewed from the standpoint of the research surrounding multimodal displays and advanced auditory cues. Previous work has found support for multimodal cues and advanced auditory cues increasing performance on the secondary search task itself and decreasing workload as compared to a visuals only interface (Chisholm et al., 2008; Jeon, Davison, Nees, Wilson, & Walker, 2009; Jeon et al., 2012; Liu, 2001; Sodnik, Dicke, Tomažič, & Billinghurst, 2008). Although these results were not all supported here, the multimodal conditions were not shown to be any worse. While it is unclear why no differences were found across workload or primary task performance between conditions the perceived performance of participants does point to the multimodal displays being perceived as better after some practice.

The results for the eye tracking, which also supported previous work (Gable et al., 2013), are particularly useful. If one considers the differences in percent time off for Spindex compared to No Sound the numeric difference in change is large (10% less time for Spindex in Drive 1, and 17% less time for Spindex in Drive 2). Although these interactions are not statistically significant they are meaningful in that this amount of practice with multimodal displays could give people more time with their eyes on the road than those using visual-only cues. When considering the results of previous work discussed early, this precious time focused on the road could make a large difference in safe driving (Klauer, Dingus, Neale, Sudweeks, & Ramsey, 2006).

The visual results are also interesting in regards to the approach taken by participants in completing the search task. Due to participants in the multimodal condition of Spindex having lower visual time off the road, even during difficult driving tasks, and the qualitative feedback stating that participants in the multimodal conditions used the auditory cues more as they gained practice we can be sure that participants did indeed use the auditory cues. This result supports the idea that the advanced auditory cues in the current research avoid the pitfall of other multimodal cues of taking too long to complete a task and could increase the usage rate of the auditory cues than a non advanced auditory cue (Brumby, Davies, Janssen, & Grace, 2011; Ranney, Mazzae, Garrott, & Goodman, 2000).

4.2.2 Practice with Advanced Auditory Cues and Multimodal Displays

While improvements were made with the displays over time, and changes in strategies took place, it was not clear if participants reached automaticity as described in Lewandowsky and Thomas (2009). Knowing how much practice is necessary to reach automaticity is difficult and most likely outside of the reach of the current research. While it seemed that the diminishing returns expected via the law of practice (Newell & Rosenbloom, 1981) had begun to be seen in the pilot, similar slowdowns of improvement were not as prevalent in the full study.

Some of the changes that were seen with practice were quite interesting though, with the difference found between the perceived helpfulness of the displays being one example. Participants across all of the conditions stated the display was more helpful at the beginning of the study than at the end. However, the participants in the Spindex and Spearcon conditions had significant interactions for this perceived helpfulness across sessions. The fact that participants in the Spindex condition perceived the displays as more helpful in Sessions 3 and 4 than 1 or 2, while the Spearcon condition participants did not rate theirs as better until Session 4, suggests that participants needed more time to adjust to the Spearcon display. The additional difference in physical workload between these cues, with Spindex being rated as having lower physical workload also point at differences between the displays, with Spindex potentially being easier to use, at least initially. Interestingly, no differences were found in actual performance with the cues, so this perceived helpfulness did not match up to any actual performance differences, which have been seen previously. This result does however, suggest that Spindex cues may be easier to learn and get used to than Spearcons.

On a related note, the low rate of improvement over time with the Spindex and Spearcon cues, as compared to the rate of improvement for the visual display suggest a few things. First these results suggest that participants were still getting used to the visual display over the course of the study. This may mean that participants are not as quick to get to a level of low returns from practice with any visual display, but that instead it may take longer to get accustomed to an interface than was expected here. This would suggest less transfer of everyday visual interface practice to a novel visual display than was expected. Second it might suggest that it did not take long for participants to get used to using the advanced auditory cues. This points to support for previous work showing the advanced auditory cues are easier to learn than other types of multimodal displays, which should be considered in the application of the current work and in how the results are considered with MRT (Dingler, Lindsay, & Walker, 2008; Jeon & Walker, 2011).

4.2.3 Multiple Resources Theory

The results here have potential theoretical implications for multitasking, particularly for Multiple Resources Theory (MRT) (Wickens, 2002; Wickens, 2008). If the study had pointed to large increases in performance for the multimodal displays over this longer period of practice than is normally done in studies with multimodal displays, then it may have been suggested that many researcher's interpretation and use of MRT has been incorrect. Such results would have suggested that results seen in previous work were not complete and suggested future work requires additional practice with multimodal displays to see their full effects on performance. This would have also suggested that the discussion by Wickens in one of his papers (2002) regarding the potential influence of practice on performance with a multimodal display for multimodal tasks holds true with the amount of practice seen here.

However, the results of this study did not show this potential result, and instead suggests that either participants need no long-term practice to see the effects of MRT, or that the full effects of MRT require longer than 100 hours of practice to show up. While this study suggests that MRT holds in this short period of time, the extent of how much applying the theory of MRT can assist in improving multi-tasking is still not clear, nor is it known how much practice is needed to determine what this advantage might be for a dual-task with multimodal displays. However, this result should considered carefully, as the lack of differences over time for the multimodal displays could be due to the well-designed nature of the advanced auditory cues being used in the present study. With previous work showing that these cues are easy to learn due to being so similar to speech (Dingler, Lindsay, & Walker, 2008; Jeon & Walker, 2011), and the differences in visual time with eyes on the road showing up without any real practice in the present study, it suggests that these cues may be very quick to learn how to use effectively. It may be that the improvements seen in the results here were simply the effect of practicing the search task itself, not in learning how to employ the auditory cues, which would explain why there were so few differences between the conditions. This would suggest that such results may not be seen with multimodal displays that are not as intuitive as advanced auditory and require more training to see this type of performance.

4.2.4 Applying Multimodal Cues

This study points to a number of factors that must be considered going forward with work involving advanced auditory cues and multimodal displays. The results suggest equal practice effects on the tasks and displays used here for this period of training. This may be because of the interface being novel to all users, since the visual task was not exactly like ones that people have used before. However, more likely is the explanation that the rates of learning across the devices and the modalities of interaction are not that different here. This suggests that MRT holds for well-designed displays without extensive practice being necessary and that previous work in the space with little practice using these displays holds up. The initial differences seen across sessions seemed to take place during the first few sessions, so it would be advisable to ensure participants have some practice, but it may be that more training than was done here is needed to see any more extensive effects of practice than have been seen initially in this type of work.

Unfortunately, the exact amount of time needed to see a more extreme effect of the ubiquitous law of practice taking place cannot be pointed to here (Newell & Rosenbloom, 1981). One must also consider a realistic time-table for a study to show the efficacy of a multimodal display, and suggesting that researchers ensure participants get over 100 minutes of practice with a display before testing differences is most likely unrealistic for many multimodal displays. Instead the results suggest that ensuring a multimodal display is intuitive and quick to learn, such as the advanced auditory cues, may help to ensure that some effect of MRT is seen fairly quickly in a study.

These results also have implications for real-world adoption of such displays. With these results pointing to participants having somewhat equal performance on the driving and search task over time, and only showing differences of visual attention between the No Sound and multimodal conditions, this may mean it could be hard to get users to adopt a multimodal display in place of a visual only display unless they are forced to do so, even with an intuitive cue such as Spindex. The initial annoyance levels of the multimodal displays, particularly of the Spearcon display also point to the potential issue in attempting to gain adoption of the displays. It may be that with more practice than gained here users would begin to see additional benefits of using the auditory displays, but it may be they do not get that far and revert back to a visual only display. One potential positive seen here in regards to the adoption of multimodal displays over a primarily visual display is that of the perceived helpfulness increasing significantly after a few sessions for the multimodal cues, and in particular the Spindex cue. This difference may mean that after a period of time participants would begin to feel more effective with the displays than they initially did, and since this effect was not seen individually for the

No Sound condition it suggests the multimodal displays have a leg up in regards to perceived helpfulness after some level of training.

4.2.5 New Questions

While some differences existed statistically and potentially meaningfully, the current study suggests that after an initial, short practice, the advantages of using well-designed multimodal displays may not be different than longer-term practice with those displays as compared to a visuals only display. There are however, a number of questions that the research brought up, which should be investigated further. First, the question of whether a difference between conditions would be seen after even longer practice is of interest. While that long of practice may not be reasonable for researchers to use in most multimodal display studies, the theoretical and potential real-world implications are still of interest. Another interesting piece was the lack of transfer of performance across some of the training to the driving situation. The decrease of search-task performance in the driving blocks in the last session as compared to the first drive is particularly interesting, and while most likely attributable to willingness to engage and the focus on the primary task more so than the secondary task, more research would be interesting to try and determine why this occurred.

It should be noted that both of these questions, and potentially other results could have been confounded by the increasing difficulty of the primary task in the training blocks. This increasing difficulty may have driven the participant to focus too much on the primary task or may have not allowed for them to have the ability to add any attention to the secondary causing a floor effect. However, it seems that a floor effect was not the case due to increasing performance on many measures including search and primary task performance in the training blocks. Instead support seems to be present for a change in willingness to engage, as participants reported changing behaviors in the subjective feedback regarding their approach to the task. These changes in approach suggest that participants simply changed their willingness to engage with the secondary task to ensure good performance on the primary task, both in the training and driving blocks and that while no differences were seen in the driving performance, this might be explained by a lack of practice at that task and a lack of transfer in learning between the ball drop and driving tasks.

4.2.6 Considerations for Future Work

In looking back at the current study and the results there were a number of pieces in that if slightly changed, may have drastically changed the outcome of the research. As a final part of the discussion, in an effort to help those in this space in the future and to make note of these potential changes the current section candidly discusses these pieces. The largest of these potential changes that should be considered is a varying of the list and/or list size when preforming research such as this. In the current work the participants reported trying to memorize the list and figuring out when the look back at the list as time went on. If the list length was changed, or if the list itself was swapped out intermittently this may have potential shown a stronger effect for those in the advanced auditory cue conditions, particularly those in the Spindex condition. Those in the No Sound condition would not have been able to memorize the list length and therefore their performance may not have improved as much as it did, and those in the Spearcon condition would have had to learn new Spearcons each time the list was changed. This may have then shown a much different result and show the true learning effects of using the advanced auditory cues instead of a combination of that and the list.

An additional piece to be considered in future work is related to the driving task. For one, switching the third drive to a more difficult, curvy road may have made the study more difficult to interpret. If it had stayed the same as the other two drives this may have allowed participants to get a better idea of the task and perform better in that last drive, or at least be more directly comparable to the other drives. On a related note, in the driving blocks the researcher did notice a few times that the drivers in the No Sound condition would be close or just over the line in more instances that those in other conditions. Due to the long-term process of the data collection, and these events being in low numbers and lasting for short periods of time, it does not show quantitatively that these events occurred. Being able to measure or somehow aim a study at investigating these differences could be of use in future work.

CHAPTER 5: CONCLUSION

The present study supported previous work with advanced auditory cues regarding visual attention and some measures of workload. It also added to the knowledge base in showing that users perceive Spindex cues as helpful more quickly than Spearcon cues. However, the more interesting results from the current study are regarding the implications on MRT and the application of multimodal displays. The results suggest that the differences seen between well-designed multimodal displays such as advanced auditory cues and visual-only displays after a short amount of practice may be no different between the displays after over 100 minutes of practice. As said before, this must be considered in the context of these advanced auditory cues of Spindex and Spearcons being well designed, and that all multimodal displays may not see the same effects. However, for such well-designed displays it may be that the results seen initially will hold for a long period of use. The meaning of these results should be considered when applying such cues to ensure that users will actually employ such displays, or when determining when one should use multimodal displays or not. Future work in the space should investigate the effects of this type of practice on less intuitive cues, and attempt to see what even longer amount of practice would do for performance with these cues.

APPENDIX A: STUDY RECRUITMENT FORM

Recruitment Description Listing for School of Psychology Subject Pool Website

We are looking for participants to be a subject for 5-30, 1-hour sessions (5-30 credits total). Participants will be asked to navigate a menu on an electronic device and may also be asked to perform a simultaneous primary task of playing a computer game or driving a virtual car. This research will help identify the most effective types of auditory menu systems and the effect of practice. Participants must have normal or corrected to normal vision, mobility, and hearing and have 2 years minimum of driving experience and a valid license. If vision corrective lenses are required they must be in the form of contacts for the study, those with glasses will not be able to participate. We also ask that participants not consume caffeine or any other stimulants and not engage in vigorous exercise for 2 hours prior to the study.

Recruitment Description Listing for word of mouth postings

We are looking for participants to be a subject for 5-30 1-hour sessions at $10 an hour ($50-300 total). Participants will be asked to navigate a menu on an electronic device and may also be asked to perform a simultaneous primary task of playing a computer game or driving a virtual car. This research will help identify the most effective types of auditory menu systems and the effect of practice. Participants must have normal or corrected to normal vision, mobility, and hearing and have 2 years minimum of driving experience and a valid license. If vision corrective lenses are required they must be in the form of contacts for the study, those with glasses will not be able to participate. We also ask that participants not consume caffeine or any other stimulants and not engage in vigorous exercise for 2 hours prior to the study.

APPENDIX B: DEMOGRAPHICS QUESTIONNAIRE

Participant Number: ________________

Demographics (once)

Age: _________________

Gender: M F

Handedness: Left Right

Native English Speaker? Yes No

Number of years with drivers license: __________________

Hand used to interact with device: Left Right

Short answer

What sort of technology do you generally use in a car?

How often do you use technology in the car?

What tasks do you use technology for in cars?

Do you have another comments or suggestions?

APPENDIX C: PREFERENCES QUESTIONNAIRE

Participant Number: ________________

Perceived performance (daily)

	☐	☐	☐	☐	☐	☐
1. How effective were you at the search task?	1	2	3	4	5	6
	Not at all Effective					Very Effective
2. How effective was this display for the task?	1	2	3	4	5	6
	Not at all Effective					Very Effective
3. How functionally helpful was this display?	1	2	3	4	5	6
	Not at all Helpful					Very Helpful
4. How annoying was this display (i.e. visual or audio + visual as applicable)?	1	2	3	4	5	6
	Not at all Annoying					Very Annoying

Perceived performance on the primary task

	☐	☐	☐	☐	☐	☐
5. How effective were you at the primary task?	1	2	3	4	5	6
	Not at all Effective					Very Effective

APPENDIX D: SCREENING FORM

<u>**Instructions:**</u> Please read the list of physical limitations, disorders, drug use, and other miscellaneous factors and determine if you qualify for this study. If you cannot truthfully answer no to all of these statements please tell the experimenter you do not qualify for this study or you do not want to participate. You **DO NOT** need to tell the experimenter which of these factors excludes you, you should simply state you do not qualify.

Statements

- I have had a valid United States driver's license for at least 2 years prior to this date.

- I have normal or corrected to normal vision and if I have vision correction I am wearing contacts or do not need them for regular driving and cell phone use (you may not wear eyeglasses).

- I have normal or corrected to normal hearing.

- I have no mobility impairments that decrease my ability to drive.

- I have not performed any vigorous exercise in the past two hours.

- I have not consumed caffeine, nicotine, or any other stimulant in the past two hours.

APPENDIX E: CONSENT FORM

Georgia Institute of Technology

Project Title: Comparison of Auditory Representations for User Menu Navigation

Investigators: Bruce Walker, Ph.D. (PI); Thom Gable (Experimenter)

Research Consent Form

You are being asked to be a participant in a research study.

Purpose:

You are being asked to be a participant in a psychology research project. Your participation will help advance scientific knowledge in the area of auditory interfaces with benefits to assistive technologies and mobile computer interfaces. Your participation will also provide you with experience conducting research in psychology.

Exclusion/ Inclusion Criteria

To participate in this study you must have normal or corrected to normal vision, mobility, and hearing and have 2 years minimum of driving experience and a valid license. If vision corrective lenses are required they must be in the form of contacts for the study, those with glasses will not be able to participate. We also ask that participants not consume caffeine or any other stimulants and not engage in vigorous exercise for 2 hours prior to the study.

Procedures:

If you decide to participate in this experiment, you will be asked to complete 5-30 1-hour sessions. The exact number of session will be determined with you based on a number of factors.

You will use a desktop computer or a mobile device such as a cell phone provided by the researcher to find several target items in an interactive menu using various cues and give feedback on preferences for different types of menus. You may also be asked to simultaneously perform another task of playing a video game or controlling a simulated automobile. The experimenter will show participants how to navigate the list on the device. The target item name will be presented on the screen and/or auditorily from the speakers simultaneously. After seeing/hearing the target item, participants will navigate the list of names by swiping an onscreen area; or gliding a finger across the list area according to the allotted condition. After the selection of the target, there will be another randomly selected target item presented. After each block, you may be asked to complete an electronic version of the NASA-TLX questionnaire (workload assessment) and may be asked to answer questions about the block. Finally, after completing all blocks and sessions, participants will fill out a short questionnaire for demographic information, indicate their preferences, and provide other comments on the. During the study an eye-tracking system and a physiological sensing system will be used to track where you are looking and track your heart rate. This will only be used during the experimental conditions and will not be used during other times in the study.

You may withdraw from participation in this study at any time.

Risks/Discomforts

Participants assigned experimental conditions involving a driving simulator may experience nausea (e.g. "simulator sickness"). Participants that experience discomfort may immediately remove themselves from the study at no penalty and will receive full credit or pay for participation. In addition, the biometric system connectors contain nickel and may cause some irritation in those with a nickel allergy. The risks posed by other activities in this study are no greater than those involved in daily activities such as watching a film or typing on the computer.

Benefits

You are not likely to benefit in any way from joining this study. However, participation in this study will contribute to our understanding of the suitability of different types of auditory interfaces for navigating menus on computing devices.

Compensation to You

A total of 1 credit hour in a GA Tech psychology course may be awarded for each session in this study via the School of Psychology Subject Pool website (total 5-30 credits across 5-30 sessions) or a participant may be paid $10 for each hour-long session (totaling $50-300 for the 5-30 sessions). If you withdraw early from a session, you will still receive the full credit or pay from that session.

"U.S. Tax Law requires that a 1099-misc be issued if U.S. tax residents receive $600 or more per calendar year. If non-U.S. tax residents receive more than $75, mandatory 30% withholding is required. Your address and citizenship/visa status may be collected for compensation purposes only. This information will be shared only with the Georgia Tech department that issues compensation, if any, for your participation."

Confidentiality

The data collected about you will be kept private to the extent allowed by law. To protect your privacy, your records will be kept under a code number rather than by name. All experimental data will be stored in a secure, locked location for the duration of the study. Your records will be kept in locked files and only study staff will be allowed to look at them. Your name and any other fact that might point to you will not appear when results of this study are presented or published. Your privacy will be protected to the extent allowed by law. To make sure that this research is being carried out in the proper way, the Georgia Institute of Technology IRB may review study records. The Office of Human Research Protections may also look over study records during required reviews.

Costs to You

There are no costs to you, except for your time.

In Case of Injury/Harm

If you are injured as a result of being in this study, please contact Bruce Walker, Ph.D. by phone at (404) 894-8265. Neither the Investigators nor Georgia Institute of Technology have made provision for payment of costs associated with any injury resulting from this study.

Participant Rights

- Your participation in this study is voluntary. You do not have to be in this study if you don't want to be.
- You have the right to change your mind and leave the study at any time without giving any reason and without penalty.
- Any new information that may make you change your mind about being in this study will be given to you.
- You will be given a copy of this consent form to keep.
- You do not waive any of your legal rights by signing this consent form.

Questions about the Study or Your Rights as a Research Participant

If you have any questions about the study, you may contact Bruce Walker Ph.D. by phone at (404) 894-8265 or by email at bruce.walker@psych.gatech.edu

If you have any questions about your rights as a research participant, you may contact Ms. Kelly Winn by phone at (404) 385-2175 or by email at Kelly.winn@gtrc.gatech.edu

If you sign below, it means that you have read (or have had read to you) the information given in this consent form, and you would like to be a volunteer in this study.

Please note that you may choose to stop participating at any time, with no penalty whatsoever.

____

Participant Name

____

Participant Signature **Date**

____

Signature of Person Obtaining Consent **Date**

APPENDIX F: PARTICIPANT INSTRUCTIONS

Overall Project

This experiment is a part of the auditory menus project, which is intended to enhance the use of electronic menus with auditory cues.

Purpose of Experiment

This research is comparing various auditory cues to improve the performance of menu navigation (secondary task) and a simple driving-like or driving task (primary task).

Procedure

In this experiment, you will be asked to navigate through a list on a device to find a requested target-items (item names) while you are playing a video game or using a driving simulator.

Video Game

You may be asked to play a video game where your goal is to catch the falling balls with the platform that you control via the keyboard. You move left and right with the arrow keys and try to catch all of the balls.

Driving Simulation

You may also be asked to perform a driving task. To complete the task correctly your goal is to stay in the lane that you start in and follow the experimenter's instructions regarding where to go. You need to try and stay in the lane you are in to the best of your ability with the steering wheel while avoiding any collisions.

Menu Navigation

You will be searching for an item on a device during these tasks. You will be able to use the device with whatever hand you choose but for the rest of the experiment you must use that same hand and control the game or drive with the other hand. Once the experiment begins you will hear a target item randomly generated from the device. Upon hearing this, you will start to navigate the list on the device by either flicking through it or using a dragging motion on the alphabetized index. The list is alphabetized by the first letter of the item. You will be asked to move down the list as quickly as possible until you find the target item. Once you find the target, clicking on the item will signal that you have reached your final destination. Make sure that the target name disappears on the screen. Unlike the functionality of a typical device, the list will not wrap around after reaching the top or end of the list. You do not have to listen to the complete name before moving to the next item if there is audio, instead, we would like you to move as quickly as possible through the list but without sacrificing accuracy. If you are sure an item is not your target item, feel free to move over that item as you navigate.

After the selection of the target, you can continue the other task alone again. After a few seconds, you will hear a new target name, and the remaining procedure is the same as the previous case.

We will give you few minutes to familiarize yourself with this task before we start the conditions.

During the study you will be asked to do each task by itself, or at the same time. Please do both tasks to the best of your ability.

In this experiment, driving task is the primary task, so you should focus on it as 80% of your resources. You can allocate your resources as 20% to your navigation task.

Between each condition you may be asked to answer some questions regarding your workload and at the end of the study you will be asked to complete a questionnaire with demographics and other information.

You will also be asked to use eye-trackers during this study. We will fit you with them at the beginning of the study and they will be used during each condition (They will not be recording when you are filling out preference and workload questionnaires). We ask that you view the driving simulator and secondary task as you normally would.

We will also ask to place a heart rate measurement device to you to track your heart rate during the study. We will go over this part of the instructions when we are ready to fit you with the device.

If you have any questions about the task you are being asked to perform, please ask the experimenter now.

If you change your mind to participate in this study, you can stop at anytime.

Now, let’s get started.

APPENDIX G: HEART RATE MONITOR PLACEMENT INSTRUCTIONS

Instructions for Applying Heart Rate Monitoring Leads

1) Using the included diagram, identify the 3 locations where the electrode pads will be applied. Two will be located just under the collarbones, preferably in the gap between the shoulder muscles. The other will be near the stomach over the bottom rib bone.

2) Gently wipe the 3 areas with a cotton swab to clear any dead skin.

3) Using an alcohol wipe, clean the 3 areas thoroughly and then let them dry to allow the electrode pads to stick cleanly.

4) Once the wiped areas are dry, place one pad in the center of each cleaned area. All pads are the same, and it does not matter which pads goes on which of the 3 areas.

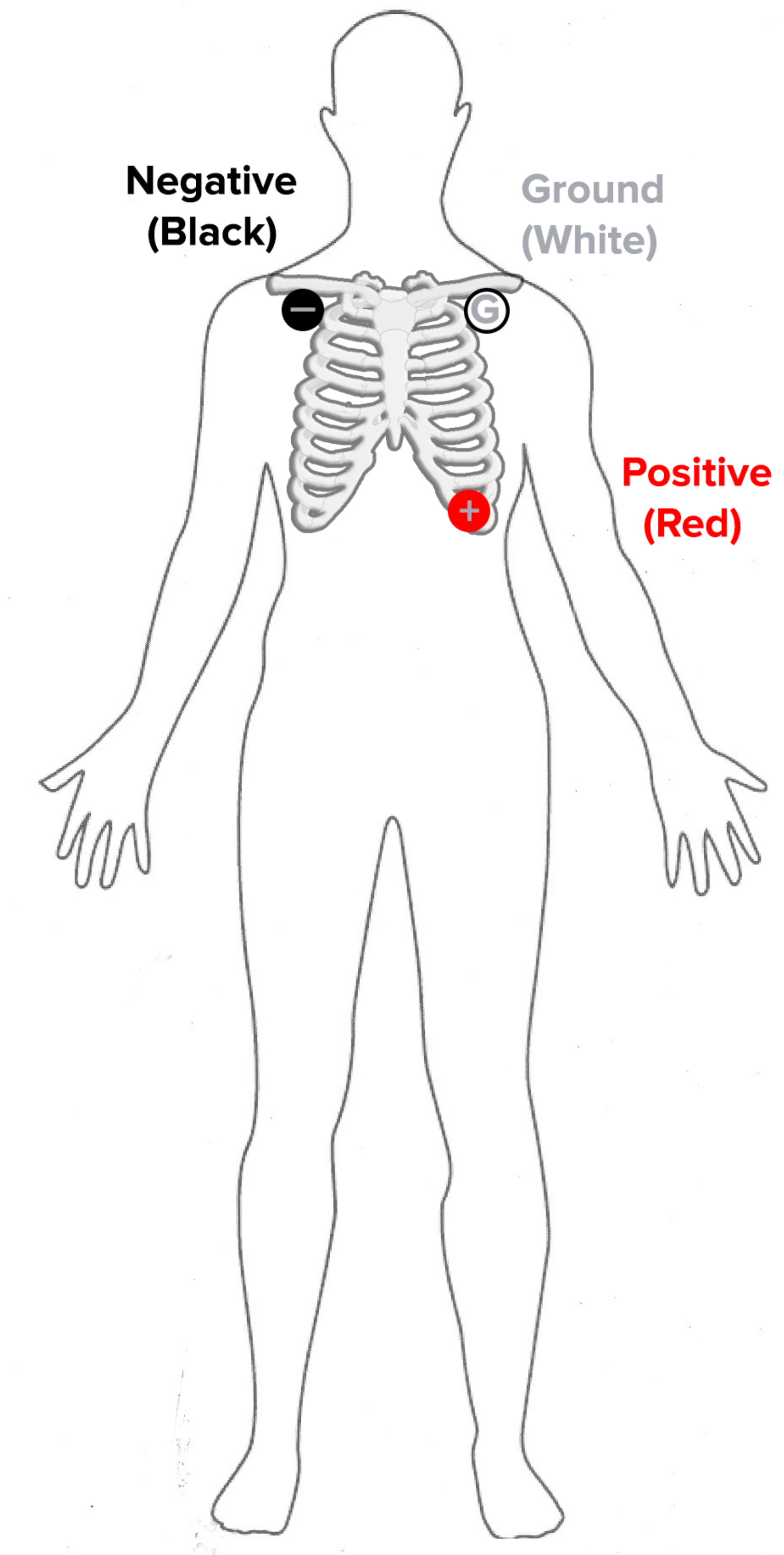
5) Grab the three wire leads for the heart rate system. They are labeled with twist ties as:

+	-	G
(Positive)	(Negative)	(Ground)

6) Snap each lead onto the proper electrode pad according to the included diagram. The leads should snap in with only a small amount of force.

7) Verify the location of the three electrodes. They should be free from all fabric, belts, and clothing and should not fall or peel off as you move. The leads should go under your clothing and out of the area by your belt.

8) Notify the experimenter that you are ready to continue.


Negative
(Black)
Ground
(White)
Positive
(Red)
G
+

APPENDIX H: DEBRIEF FORM

Thanks and Introduction

First of all, thank you for your participation in this experiment with the Sonification Lab in school of psychology.

Overall Project

This experiment is a part of the auditory menus project, which is intended to enhance the use of electronic menus with auditory cues.

Purpose of Experiment

The purpose of this experiment is to analyze different types of auditory menus and determine which features are most useful.

Revelation of Experiment Condition

The purpose of this experiment is to compare cognitive load, visual behaviors, and primary task performance when learning how to interact with a secondary task (song searching). Auditory cues are believed to improve the performance of menu navigation (secondary task) and primary task (ball catching or driving) while reducing the cognitive effort of users and this study aims to look at the effect of experience on this factor.

Meaning of Expected Results

By analyzing your performance on the tasks (time to find the target, accuracy, etc.) along with many other participants in this study, we can make generalizations about how experience effects how people interact with auditory menus and how this affects their performance on that and other tasks.

Confidentiality and Anonymity

The results of your experiment will be used only for psychological study and never used for any other purposes. The data that is collected from you will be kept private to the extent allowed by law. To protect your privacy, your records will be kept under a code number rather than by name. Your records will be kept in locked files and only study staff will be allowed to look at them. Your name and any other fact that might point to you will not appear when results of this study are presented or published. To make sure that this research is being carried out in the proper way, the Georgia Institute of Technology IRB will review study records. Again, your privacy will be protected to the extent allowed by law.

Conclusion

All of the experiment procedures are finished. We very much appreciate your efforts!

Contact Information

For further information of this research, contact:

Principal Investigator

Dr. Bruce Walker, Ph.D. (bruce.walker@psych.gatech.edu)

Experimenter

Thomas Gable (thomas.gable@gatech.edu)

APPENDIX I: PILOT DESCRIPTVIE TABLES

Table I.1 – Pilot Search Task Performance Data for each Participant by Training Block

Condition	No-Sound		Spindex		Spearcon		Block Mean	
Participant Number	1	4	3	5	2	6	*M*	*SD*
Mean Number Searches								
Block 1	16.00	20.00	11.00	14.00	30.00	17.00	18.00	6.60
Block 2	13.00	20.00	10.00	20.00	23.00	21.00	17.83	5.12
Block 3	18.00	16.00	11.00	21.00	26.00	22.00	19.00	5.22
Block 4	18.00	18.00	9.00	22.00	33.00	22.00	20.33	7.81
Block 5	13.00	17.00	8.00	19.00	37.00	22.00	19.33	9.93
Block 6	22.00	18.00	8.00	23.00	35.00	22.00	21.33	8.71
Block 7	25.00	16.00	10.00	20.00	32.00	15.00	19.67	7.87
Block 8	25.00	16.00	12.00	24.00	32.00	21.00	21.67	7.06
Block 9	25.00	17.00	6.00	25.00	33.00	20.00	21.00	9.14
Block 10	23.00	19.00	13.00	28.00	27.00	25.00	22.50	5.65
Block 11	27.00	18.00	11.00	19.00	31.00	25.00	21.83	7.22
Block 12	29.00	18.00	10.00	25.00	33.00	24.00	23.17	8.18
Block 13	28.00	15.00	8.00	24.00	29.00	21.00	20.83	8.08
Block 14	27.00	15.00	12.00	27.00	27.00	18.00	21.00	6.84
Block 15	33.00	13.00	13.00	24.00	32.00	16.00	21.83	9.20
Participant Mean	22.80	17.07	10.13	22.33	30.67	20.73	-	-
Participant *SD*	6.00	1.94	2.03	3.60	3.70	3.06	-	-
Percent Accuracy								
Block 1	93.75	80.00	72.73	92.86	93.33	82.35	85.84	8.79
Block 2	100.00	65.00	60.00	100.00	100.00	95.24	86.71	18.91
Block 3	100.00	93.75	81.82	100.00	96.15	90.91	93.77	6.85
Block 4	88.89	61.11	88.89	95.45	100.00	90.91	87.54	13.65
Block 5	84.62	64.71	75.00	100.00	94.59	90.91	84.97	13.15
Block 6	95.45	61.11	75.00	100.00	97.14	100.00	88.12	16.24
Block 7	92.00	68.75	90.00	90.00	96.88	93.33	88.49	10.00
Block 8	96.00	62.50	100.00	100.00	100.00	95.24	92.29	14.75
Block 9	92.00	82.35	83.33	100.00	93.94	95.00	91.10	6.93
Block 10	82.61	57.89	84.62	96.43	96.30	100.00	86.31	15.58
Block 11	92.59	66.67	72.73	89.47	100.00	100.00	86.91	14.09
Block 12	96.55	88.89	80.00	100.00	96.97	100.00	93.74	7.86
Block 13	85.71	66.67	75.00	100.00	96.55	100.00	87.32	14.05
Block 14	92.59	73.33	83.33	96.30	92.59	100.00	89.69	9.74
Block 15	81.82	84.62	100.00	100.00	96.88	93.75	92.84	7.86
Participant Mean	91.64	71.82	81.50	97.37	96.75	95.18	-	-
Participant *SD*	5.82	11.29	10.60	3.83	2.46	5.10	-	-
Mean Search Time (s)								
Block 1	28.03	43.89	13.83	25.60	26.60	30.02	28.27	9.64

Block 2	28.79	40.97	19.63	27.17	37.62	23.67	23.69	8.18
Block 3	27.99	52.74	17.41	26.60	26.13	22.99	22.08	12.24
Block 4	28.44	56.66	13.26	26.39	29.90	22.38	22.03	14.58
Block 5	30.19	60.04	10.86	29.06	34.13	25.36	21.69	16.08
Block 6	25.85	60.25	11.29	23.61	19.64	19.01	21.27	17.22
Block 7	30.19	54.23	12.80	35.85	19.97	24.17	34.15	14.50
Block 8	25.93	44.38	13.38	36.41	18.66	19.90	22.85	11.79
Block 9	24.44	39.62	13.40	33.11	18.99	18.34	23.16	9.94
Block 10	22.53	37.07	16.66	25.33	23.25	14.41	18.42	7.97
Block 11	23.51	43.89	13.89	25.77	16.55	24.32	16.61	10.53
Block 12	23.74	46.22	13.39	28.33	15.70	19.04	19.76	11.98
Block 13	26.32	49.65	15.91	33.94	17.72	19.93	20.76	12.81
Block 14	24.87	37.63	17.37	34.93	16.36	16.98	25.95	9.54
Block 15	26.43	40.28	13.18	40.95	13.45	18.76	31.95	12.65
Participant Mean	28.32	50.17	16.42	32.21	23.31	23.29	-	-
Participant *SD*	2.40	7.93	2.45	5.21	7.14	3.92	-	-

Table I.2 - Pilot Ball Drop Performance Data for each Participant by Training Block

Condition	No-Sound		Spindex		Spearcon		Block Mean	
Participant Number	1	4	3	5	2	6	*M*	*SD*
Percent Accuracy								
Block 1	60.69	58.67	50.68	55.34	59.03	53.62	56.34	3.80
Block 2	69.28	68.74	43.95	64.39	71.42	59.01	62.80	10.23
Block 3	67.74	71.40	48.32	64.20	59.39	63.42	62.41	8.01
Block 4	71.56	54.42	62.76	53.63	63.63	62.87	61.48	6.65
Block 5	74.66	61.47	61.72	54.13	64.35	66.45	63.80	6.76
Block 6	74.73	64.25	69.43	52.56	62.56	66.76	65.05	7.47
Block 7	62.46	63.34	57.44	57.41	73.93	64.36	63.16	6.06
Block 8	63.37	70.48	56.90	60.93	75.21	65.19	65.35	6.61
Block 9	66.57	71.99	59.79	59.57	78.33	62.80	66.51	7.44
Block 10	71.60	56.71	71.35	51.32	80.82	64.49	66.05	10.81
Block 11	69.51	59.56	68.11	51.55	77.62	60.67	64.50	9.13
Block 12	72.28	63.36	65.85	51.15	76.74	62.29	65.28	8.87
Block 13	64.12	67.68	58.53	62.53	76.44	62.73	65.34	6.18
Block 14	64.30	64.56	64.10	55.74	68.53	60.40	62.94	4.37
Block 15	60.54	61.95	67.44	57.49	73.75	58.04	63.20	6.28
Participant Mean	72.39	68.47	64.74	60.85	75.84	66.65	-	-
Participant *SD*	4.83	5.35	8.01	4.62	7.29	3.45	-	-
Balls Dropped								
Block 1	1306.00	981.00	1175.00	1153.00	1137.00	982.00	1122.33	124.36
Block 2	1314.00	980.00	1316.00	1145.00	1145.00	995.00	1149.17	146.58
Block 3	1310.00	980.00	1307.00	1150.00	1316.00	989.00	1175.33	160.48

Block 4	1310.00	1133.00	1264.00	1336.00	1187.00	1073.00	1217.17	103.64
Block 5	1307.00	1144.00	1282.00	1309.00	1179.00	1080.00	1216.83	96.26
Block 6	1320.00	1148.00	1255.00	1312.00	1187.00	1072.00	1215.67	97.73
Block 7	1389.00	1182.00	1312.00	1316.00	1075.00	1175.00	1241.50	116.59
Block 8	1389.00	1180.00	1312.00	1310.00	1067.00	1173.00	1238.50	118.57
Block 9	1399.00	1179.00	1310.00	1313.00	1067.00	1182.00	1241.67	120.42
Block 10	1305.00	1314.00	1311.00	1476.00	1071.00	1178.00	1275.83	137.94
Block 11	1317.00	1315.00	1315.00	1489.00	1069.00	1263.00	1294.67	134.92
Block 12	1314.00	1312.00	1311.00	1478.00	1077.00	1259.00	1291.83	128.96
Block 13	1478.00	1320.00	1414.00	1314.00	1186.00	1313.00	1337.50	100.00
Block 14	1481.00	1317.00	1416.00	1317.00	1184.00	1309.00	1337.33	101.95
Block 15	1482.00	1315.00	1423.00	1316.00	1178.00	1314.00	1338.00	105.02
Participant Mean	1458.64	1271.43	1408.79	1409.57	1223.21	1239.79	-	-
Participant *SD*	69.72	129.19	64.75	108.78	71.24	119.73	-	-

Table I.3 - Pilot Visual Behaviors Data for each Participant by Training Block

Condition	No-Sound		Spindex		Spearcon		Block Mean	
Participant Number	1	4	3	5	2	6	*M*	*SD*
Percent Time Off								
Block 1	5.40	13.00	14.05	55.37	23.69	21.53	22.17	17.53
Block 2	6.24	7.00	7.48	91.27	16.06	4.82	22.14	34.10
Block 3	6.61	2.49	11.20	92.94	18.36	NA	26.32	37.71
Block 4	12.27	14.86	2.43	89.55	19.91	9.51	24.75	32.27
Block 5	4.09	8.93	5.98	81.49	25.41	12.86	23.13	29.58
Block 6	10.83	17.88	2.77	75.32	26.57	7.91	23.55	26.68
Block 7	13.34	13.15	8.12	43.71	16.85	18.95	19.02	12.65
Block 8	15.87	13.57	5.51	20.08	16.13	11.46	13.77	4.97
Block 9	13.74	13.31	3.71	76.31	14.91	19.31	23.55	26.35
Block 10	10.23	24.89	2.34	14.78	9.13	10.04	11.90	7.52
Block 11	9.65	13.08	1.39	10.49	9.56	10.44	9.10	3.99
Block 12	13.17	11.79	2.24	11.93	7.77	2.93	8.30	4.79
Block 13	4.09	11.09	2.82	11.56	9.61	8.58	7.96	3.67
Block 14	9.49	8.92	3.35	23.08	6.57	6.51	9.65	6.93
Block 15	5.82	10.56	0.78	20.63	8.71	8.48	9.16	6.56
Participant Mean	9.39	12.30	4.94	47.90	15.28	10.95	-	-
Participant *SD*	3.82	5.02	3.81	33.43	6.62	5.52	-	-
Glance Rate (glance/min)								
Block 1	7.50	11.70	16.70	15.30	15.80	12.60	13.27	3.42
Block 2	7.60	7.60	7.20	7.20	8.00	4.40	7.00	1.31
Block 3	7.40	3.40	7.20	9.30	7.80	NA	7.02	2.18
Block 4	10.10	9.00	2.30	11.70	9.80	9.30	8.70	3.27
Block 5	4.00	7.60	2.80	10.70	8.80	7.40	6.88	2.97
Block 6	7.10	9.30	2.20	22.10	10.70	9.00	10.07	6.60

Block 7	10.50	7.40	3.60	21.40	6.90	16.30	11.02	6.65
Block 8	12.90	5.90	2.90	10.30	5.90	11.30	8.20	3.87
Block 9	10.30	6.60	3.10	14.90	5.20	12.50	8.77	4.55
Block 10	10.00	17.70	1.90	14.00	5.90	7.00	9.42	5.74
Block 11	10.90	6.10	1.80	9.90	6.40	7.10	7.03	3.22
Block 12	12.30	5.30	1.60	9.90	4.70	4.00	6.30	4.00
Block 13	5.10	6.50	2.10	7.90	4.80	4.90	5.22	1.94
Block 14	9.80	5.50	2.90	10.40	7.90	4.10	6.77	3.08
Block 15	6.50	6.20	1.00	8.30	6.10	4.40	5.42	2.49
Participant Mean	8.80	7.72	3.95	12.22	7.65	8.16	-	-
Participant *SD*	2.57	3.38	3.97	4.55	2.87	3.85	-	-
Mean Dwell Length (ms)								
Block 1	431.78	666.67	504.69	2171.35	899.68	1025.26	949.91	639.85
Block 2	492.98	552.41	623.38	7605.56	1204.17	657.20	1855.95	2828.28
Block 3	536.26	439.71	933.56	5996.42	1412.18	NA	1863.63	2341.91
Block 4	728.71	990.56	633.33	4592.31	1218.71	613.80	1462.90	1550.75
Block 5	612.92	704.61	1280.95	4569.78	1732.77	1042.57	1657.26	1483.65
Block 6	915.49	1153.58	754.55	2044.95	1489.88	527.04	1147.58	550.61
Block 7	762.06	1065.99	1353.24	1225.55	1465.22	697.55	1094.93	313.19
Block 8	737.98	1379.94	1140.23	1169.58	1639.83	608.41	1112.66	386.78
Block 9	800.16	1209.85	718.82	3072.93	1720.19	927.07	1408.17	892.69
Block 10	613.67	843.88	738.60	633.33	928.81	860.48	769.79	128.82
Block 11	531.35	1286.61	463.89	636.03	896.09	882.16	782.69	304.31
Block 12	642.41	1334.28	840.63	723.23	991.49	438.75	828.46	309.88
Block 13	480.72	1024.10	804.76	878.06	1201.04	1050.00	906.45	250.45
Block 14	581.29	973.03	692.53	1331.41	499.16	953.25	838.44	308.82
Block 15	536.92	1022.04	470.00	1491.16	856.56	1156.44	922.19	386.21
Participant Mean	626.98	976.48	796.88	2542.78	1210.38	817.14	-	-
Participant *SD*	136.22	285.46	275.75	2172.77	363.98	224.01	-	-

Table I.4 - Pilot Heart Rate Percent Difference Data for each Participant by Training Block

Condition	No-Sound		Spindex		Spearcon		Block Mean	
Participant Number	1	4	3	5	2	6	*M*	*SD*
Percent Difference HR								
Block 1	84.74	92.06	85.19	57.45	77.76	81.20	79.73	11.91
Block 2	84.59	84.86	81.50	57.25	77.42	79.19	77.47	10.33
Block 3	84.32	84.14	82.11	58.67	73.12	78.42	76.80	9.83
Block 4	79.33	96.49	89.73	67.98	72.76	76.59	80.48	10.72
Block 5	79.20	94.86	93.28	66.95	72.06	76.10	80.41	11.36
Block 6	80.58	92.64	89.43	67.21	72.66	73.73	79.38	10.03
Block 7	81.62	94.77	86.31	68.25	71.65	74.15	79.46	10.01

Block 8	78.58	94.33	85.25	68.37	71.96	73.97	78.74	9.61
Block 9	77.09	94.47	83.89	68.74	70.54	72.15	77.81	9.82
Block 10	73.25	82.45	86.83	73.62	64.88	NA	76.21	8.60
Block 11	74.71	81.53	86.32	70.59	66.27	NA	75.88	8.10
Block 12	70.72	82.39	86.77	70.65	64.78	NA	75.06	9.15
Block 13	63.63	84.64	93.17	65.66	74.33	68.00	74.91	11.74
Block 14	63.35	79.36	91.90	66.15	72.85	66.81	73.40	10.72
Block 15	63.74	78.19	92.81	65.19	79.79	66.44	74.36	11.35
Participant Mean	75.96	87.81	87.63	66.18	72.19	73.90	-	-
Participant *SD*	7.57	6.52	3.91	4.85	4.38	4.84	-	-
Percent Difference HRV								
Block 1	71.18	70.08	67.29	42.92	73.67	57.06	63.70	11.70
Block 2	77.60	64.39	64.32	47.17	76.75	51.21	63.57	12.60
Block 3	71.84	64.22	66.40	49.71	83.41	52.38	64.66	12.50
Block 4	70.70	82.97	62.44	50.61	73.96	44.25	64.16	14.67
Block 5	68.19	80.56	69.40	43.02	78.12	50.75	65.01	15.03
Block 6	75.14	83.30	67.50	55.74	74.37	50.82	67.81	12.42
Block 7	72.64	73.83	66.57	55.01	70.58	47.77	64.40	10.62
Block 8	73.35	74.45	65.94	54.19	76.92	51.78	66.11	10.83
Block 9	64.85	72.65	57.64	55.09	73.53	55.49	63.21	8.42
Block 10	68.25	60.84	71.35	49.02	69.09	NA	63.71	9.11
Block 11	67.82	56.04	62.79	42.22	65.59	NA	58.89	10.32
Block 12	64.45	68.17	66.87	44.63	64.90	NA	61.80	9.72
Block 13	71.18	70.96	60.32	37.97	71.51	60.03	62.00	12.95
Block 14	74.96	61.91	66.20	39.91	73.49	60.88	62.89	12.67
Block 15	70.70	58.69	61.52	50.69	49.92	66.23	59.63	8.30
Participant Mean	70.86	69.54	65.10	47.86	71.72	54.05	-	-
Participant *SD*	3.70	8.60	3.60	5.84	7.65	6.14	-	-

Table I.5 - Pilot NASA-TLX Workload Data for each Participant by Training Block

Condition	No-Sound		Spindex		Spearcon		Block Mean	
Participant Number	1	4	3	5	2	6	*M*	*SD*
Mental Workload								
Block 1	90.00	95.00	95.00	90.00	90.00	90.00	91.67	2.58
Block 2	90.00	95.00	100.00	85.00	90.00	90.00	91.67	5.16
Block 3	85.00	95.00	100.00	75.00	90.00	90.00	89.17	8.61
Block 4	90.00	95.00	100.00	80.00	90.00	80.00	89.17	8.01
Block 5	90.00	90.00	100.00	75.00	65.00	90.00	85.00	12.65
Block 6	90.00	95.00	100.00	75.00	75.00	80.00	85.83	10.68
Block 7	90.00	85.00	100.00	80.00	75.00	85.00	85.83	8.61
Block 8	85.00	80.00	100.00	75.00	60.00	90.00	81.67	13.66
Block 9	85.00	85.00	100.00	80.00	70.00	95.00	85.83	10.68
Block 10	85.00	90.00	100.00	75.00	50.00	90.00	81.67	17.51
Block 11	85.00	90.00	100.00	80.00	45.00	90.00	81.67	19.15

Block 12	85.00	90.00	100.00	75.00	45.00	90.00	80.83	19.34
Block 13	90.00	85.00	100.00	70.00	75.00	90.00	85.00	10.95
Block 14	85.00	85.00	95.00	75.00	70.00	90.00	83.33	9.31
Block 15	85.00	90.00	100.00	75.00	80.00	90.00	86.67	8.76
Participant Mean	87.33	89.67	99.33	77.67	71.33	88.67	-	-
Participant *SD*	2.58	4.81	1.76	4.95	15.86	3.99	-	-
Physical Workload								
Block 1	70.00	25.00	100.00	20.00	75.00	20.00	51.67	34.45
Block 2	80.00	25.00	80.00	25.00	90.00	15.00	52.50	34.17
Block 3	85.00	15.00	85.00	25.00	85.00	20.00	52.50	35.74
Block 4	85.00	15.00	70.00	25.00	75.00	10.00	46.67	33.57
Block 5	85.00	15.00	70.00	35.00	65.00	15.00	47.50	29.96
Block 6	85.00	20.00	85.00	25.00	70.00	15.00	50.00	33.47
Block 7	85.00	10.00	60.00	30.00	55.00	15.00	42.50	29.11
Block 8	85.00	10.00	60.00	30.00	50.00	15.00	41.67	28.75
Block 9	85.00	10.00	70.00	25.00	65.00	15.00	45.00	32.09
Block 10	80.00	15.00	75.00	20.00	50.00	15.00	42.50	30.12
Block 11	80.00	20.00	55.00	25.00	45.00	15.00	40.00	24.90
Block 12	85.00	15.00	45.00	20.00	40.00	15.00	36.67	26.96
Block 13	80.00	10.00	35.00	20.00	55.00	15.00	35.83	27.10
Block 14	80.00	15.00	45.00	25.00	70.00	15.00	41.67	28.23
Block 15	85.00	15.00	75.00	25.00	75.00	20.00	49.17	32.31
Participant Mean	82.33	15.67	67.33	25.00	64.33	15.67	-	-
Participant *SD*	4.17	4.95	17.51	4.23	14.74	2.58	-	-
Performance								
Block 1	85.00	85.00	100.00	45.00	40.00	15.00	61.67	33.12
Block 2	80.00	85.00	100.00	30.00	30.00	20.00	57.50	34.60
Block 3	75.00	75.00	90.00	35.00	65.00	15.00	59.17	28.36
Block 4	95.00	75.00	75.00	40.00	70.00	30.00	64.17	24.38
Block 5	85.00	95.00	65.00	45.00	30.00	20.00	56.67	30.11
Block 6	75.00	95.00	70.00	35.00	35.00	35.00	57.50	26.03
Block 7	70.00	80.00	70.00	40.00	25.00	15.00	50.00	27.02
Block 8	80.00	75.00	60.00	35.00	30.00	90.00	61.67	24.63
Block 9	75.00	75.00	80.00	30.00	35.00	80.00	62.50	23.40
Block 10	75.00	90.00	50.00	30.00	25.00	65.00	55.83	25.58
Block 11	65.00	90.00	50.00	30.00	25.00	70.00	55.00	24.90
Block 12	60.00	85.00	35.00	30.00	20.00	40.00	45.00	23.66
Block 13	80.00	85.00	65.00	30.00	40.00	45.00	57.50	22.53
Block 14	75.00	90.00	55.00	25.00	50.00	65.00	60.00	22.36
Block 15	70.00	95.00	35.00	25.00	45.00	50.00	53.33	25.43
Participant Mean	76.33	85.00	66.67	33.67	37.67	43.67	-	-
Participant *SD*	8.55	7.56	20.32	6.40	14.62	25.32	-	-
Effort								
Block 1	90.00	90.00	100.00	90.00	85.00	85.00	90.00	5.48
Block 2	85.00	95.00	80.00	90.00	85.00	95.00	88.33	6.06
Block 3	90.00	100.00	70.00	85.00	85.00	95.00	87.50	10.37

Block 4	90.00	85.00	85.00	90.00	70.00	90.00	85.00	7.75
Block 5	90.00	90.00	70.00	80.00	70.00	90.00	81.67	9.83
Block 6	90.00	95.00	70.00	90.00	75.00	90.00	85.00	10.00
Block 7	90.00	85.00	70.00	80.00	70.00	90.00	80.83	9.17
Block 8	90.00	85.00	55.00	75.00	70.00	95.00	78.33	14.72
Block 9	90.00	55.00	75.00	70.00	80.00	90.00	76.67	13.29
Block 10	90.00	85.00	55.00	80.00	80.00	90.00	80.00	13.04
Block 11	90.00	85.00	45.00	80.00	70.00	95.00	77.50	18.10
Block 12	90.00	85.00	40.00	70.00	70.00	95.00	75.00	20.00
Block 13	90.00	85.00	80.00	75.00	75.00	95.00	83.33	8.16
Block 14	85.00	65.00	70.00	70.00	60.00	90.00	73.33	11.69
Block 15	90.00	75.00	75.00	70.00	60.00	90.00	76.67	11.69
Participant Mean	89.33	84.00	69.33	79.67	73.67	91.67	-	-
Participant *SD*	1.76	11.53	15.45	7.90	8.12	3.09	-	-
Temporal Workload								
Block 1	95.00	100.00	100.00	95.00	60.00	95.00	90.83	15.30
Block 2	95.00	100.00	100.00	95.00	80.00	95.00	94.17	7.36
Block 3	95.00	95.00	95.00	85.00	70.00	90.00	88.33	9.83
Block 4	95.00	100.00	95.00	90.00	60.00	90.00	88.33	14.38
Block 5	95.00	95.00	95.00	85.00	55.00	90.00	85.83	15.63
Block 6	95.00	95.00	95.00	85.00	55.00	90.00	85.83	15.63
Block 7	95.00	100.00	95.00	85.00	55.00	90.00	86.67	16.33
Block 8	95.00	95.00	95.00	80.00	70.00	90.00	87.50	10.37
Block 9	95.00	95.00	95.00	85.00	70.00	95.00	89.17	10.21
Block 10	95.00	90.00	90.00	85.00	60.00	95.00	85.83	13.20
Block 11	95.00	90.00	95.00	75.00	70.00	95.00	86.67	11.25
Block 12	95.00	90.00	95.00	85.00	60.00	95.00	86.67	13.66
Block 13	100.00	95.00	95.00	75.00	65.00	90.00	86.67	13.66
Block 14	95.00	95.00	90.00	80.00	65.00	95.00	86.67	12.11
Block 15	95.00	100.00	95.00	85.00	70.00	95.00	90.00	10.95
Participant Mean	95.33	95.67	95.00	84.67	64.33	92.67	-	-
Participant *SD*	1.29	3.72	2.67	5.81	7.29	2.58	-	-
Frustration								
Block 1	95.00	85.00	100.00	65.00	65.00	75.00	80.83	14.97
Block 2	95.00	85.00	90.00	70.00	40.00	95.00	79.17	21.31
Block 3	95.00	85.00	90.00	75.00	45.00	100.00	81.67	19.92
Block 4	90.00	75.00	70.00	70.00	40.00	75.00	70.00	16.43
Block 5	85.00	85.00	70.00	65.00	20.00	85.00	68.33	25.23
Block 6	85.00	90.00	75.00	70.00	25.00	65.00	68.33	23.17
Block 7	90.00	75.00	85.00	65.00	20.00	75.00	68.33	25.23
Block 8	80.00	55.00	80.00	60.00	20.00	95.00	65.00	26.46
Block 9	85.00	55.00	85.00	45.00	55.00	80.00	67.50	17.82
Block 10	80.00	75.00	60.00	35.00	55.00	55.00	60.00	16.12
Block 11	85.00	75.00	70.00	25.00	35.00	70.00	60.00	24.08
Block 12	80.00	75.00	65.00	35.00	25.00	50.00	55.00	22.14
Block 13	85.00	70.00	80.00	20.00	20.00	40.00	52.50	29.62

Block 14	85.00	55.00	70.00	25.00	20.00	65.00	53.33	25.82
Block 15	80.00	50.00	40.00	30.00	20.00	75.00	49.17	24.17
Participant Mean	86.33	72.67	75.33	50.33	33.67	73.33	-	-
Participant *SD*	5.50	13.07	14.57	20.04	15.52	16.87	-	-
Composite Workload								
Block 1	91.67	92.67	98.33	84.00	73.00	77.00	86.11	9.82
Block 2	91.33	94.33	94.00	80.33	83.33	89.33	88.78	5.76
Block 3	92.00	93.00	92.33	71.00	80.33	89.33	86.33	8.86
Block 4	91.00	90.33	88.00	73.00	76.67	81.33	83.39	7.55
Block 5	90.67	91.00	86.00	71.67	56.00	84.00	79.89	13.65
Block 6	90.67	94.67	88.33	64.67	59.67	80.33	79.72	14.47
Block 7	91.00	88.67	89.67	73.00	52.00	81.00	79.22	14.97
Block 8	89.33	83.67	86.00	64.67	54.00	92.00	78.28	15.32
Block 9	89.67	74.00	91.00	69.00	60.00	90.67	79.06	13.27
Block 10	88.67	87.67	79.00	62.33	51.67	85.33	75.78	15.31
Block 11	89.00	87.67	82.00	68.00	50.00	88.67	77.56	15.67
Block 12	89.33	87.33	79.33	67.00	45.33	83.67	75.33	16.70
Block 13	92.00	87.33	89.67	64.00	63.00	81.67	79.61	12.95
Block 14	87.67	83.67	80.67	62.67	62.33	86.67	77.28	11.71
Block 15	89.33	89.67	86.00	67.33	64.00	80.33	79.44	11.24
Participant Mean	90.22	88.38	87.36	69.51	62.09	84.76	-	-
Participant *SD*	1.32	5.22	5.53	6.25	11.56	4.51	-	-

Table I.6 - Pilot Survey Response Values for each Participant by Session

Condition	No-Sound		Spindex		Spearcon		Block Mean	
Participant Number	1	4	3	5	2	6	*M*	*SD*
Effective at Search								
Session 1	2.00	2.00	2.00	4.00	4.00	3.00	2.83	0.98
Session 2	3.00	2.00	5.00	4.00	5.00	4.00	3.83	1.17
Session 3	2.00	3.00	4.00	5.00	5.00	4.00	3.83	1.17
Session 4	3.00	3.00	5.00	4.00	5.00	6.00	4.33	1.21
Session 5	4.00	2.00	5.00	4.00	4.00	4.00	3.83	0.98
Participant Mean	2.80	2.40	4.20	4.20	4.60	4.20	-	-
Participant *SD*	0.84	0.55	1.30	0.45	0.55	1.10	-	-
Display is Effective								
Session 1	2.00	2.00	3.00	4.00	4.00	4.00	3.17	0.98
Session 2	2.00	2.00	4.00	4.00	4.00	4.00	3.33	1.03
Session 3	2.00	2.00	4.00	4.00	4.00	4.00	3.33	1.03
Session 4	2.00	2.00	4.00	4.00	4.00	5.00	3.50	1.22
Session 5	2.00	3.00	4.00	4.00	4.00	5.00	3.67	1.03
Participant Mean	2.00	2.20	3.80	4.00	4.00	4.40	-	-
Participant *SD*	0.00	0.45	0.45	0.00	0.00	0.55	-	-
Display is Helpful								
Session 1	2.00	2.00	4.00	4.00	4.00	4.00	3.33	1.03

Session 2	2.00	2.00	4.00	5.00	3.00	4.00	3.33	1.21
Session 3	2.00	2.00	3.00	4.00	4.00	5.00	3.33	1.21
Session 4	2.00	2.00	4.00	4.00	4.00	5.00	3.50	1.22
Session 5	2.00	2.00	4.00	4.00	3.00	5.00	3.33	1.21
Participant Mean	2.00	2.00	3.80	4.20	3.60	4.60	-	-
Participant *SD*	0.00	0.00	0.45	0.45	0.55	0.55	-	-
Display is Annoying								
Session 1	4.00	4.00	2.00	2.00	3.00	5.00	3.33	1.21
Session 2	4.00	3.00	2.00	3.00	3.00	3.00	3.00	0.63
Session 3	3.00	3.00	2.00	2.00	3.00	4.00	2.83	0.75
Session 4	3.00	4.00	2.00	2.00	2.00	2.00	2.50	0.84
Session 5	3.00	4.00	2.00	3.00	2.00	2.00	2.67	0.82
Participant Mean	3.40	3.60	2.00	2.40	2.60	3.20	-	-
Participant *SD*	0.55	0.55	0.00	0.55	0.55	1.30	-	-
Effective at Primary Task								
Session 1	2.00	2.00	3.00	3.00	4.00	3.00	2.83	0.75
Session 2	2.00	2.00	3.00	4.00	3.00	4.00	3.00	0.89
Session 3	2.00	3.00	3.00	4.00	4.00	2.00	3.00	0.89
Session 4	3.00	3.00	4.00	3.00	5.00	4.00	3.67	0.82
Session 5	2.00	3.00	4.00	4.00	3.00	3.00	3.17	0.75
Participant Mean	2.20	2.60	3.40	3.60	3.80	3.20	-	-
Participant *SD*	0.45	0.55	0.55	0.55	0.84	0.84	-	-

APPENDIX J: DESCRIPTVIE TABLES

Table J.1 – Search Task Performance Descriptive Data for the Conditions by Drive

Driving Block	No-Sound M	No-Sound SD	Spindex M	Spindex SD	Spearcon M	Spearcon SD	Drive Mean M	Drive Mean SD
Mean Number Searches								
Drive 1	21.94	6.07	21.94	6.54	18.75	6.53	20.92	6.44
Drive 2	18.69	5.77	21.22	6.86	19.13	5.60	19.74	6.12
Drive 3	18.06	5.84	19.50	6.95	17.06	4.75	18.26	5.93
Condition Mean	19.56	5.48	20.89	6.13	18.31	5.20	-	-
Percent Accuracy								
Drive 1	93.34	5.30	93.99	7.87	94.48	3.70	93.98	5.81
Drive 2	96.57	4.96	95.51	4.89	95.28	4.25	95.74	4.63
Drive 3	97.36	3.41	93.74	6.07	96.10	5.28	95.63	5.25
Condition Mean	95.75	2.41	94.41	4.41	95.29	2.35	-	-
Mean Search Time (s)								
Drive 1	15.43	3.67	17.53	6.66	17.26	5.43	16.77	5.42
Drive 2	19.91	6.87	17.13	6.26	17.80	5.39	18.23	6.18
Drive 3	21.76	9.19	19.69	8.50	20.02	6.20	20.45	7.95
Condition Mean	19.03	5.80	18.11	6.26	18.36	5.01	-	-

Table J.2 - Baseline Driving Performance Values for Conditions by Baseline Drive

Baseline Driving Block	No-Sound M	No-Sound SD	Spindex M	Spindex SD	Spearcon M	Spearcon SD	Baseline Mean M	Baseline Mean SD
Mean Follow Distance (feet)								
Baseline 1	98.92	16.66	108.44	22.60	97.04	21.08	101.86	20.56
Baseline 2	82.51	12.93	89.41	16.14	83.68	21.66	85.42	17.03
Condition Mean	90.72	11.22	98.92	14.98	90.36	18.62	-	-
Follow Distance STD (feet)								
Baseline 1	42.75	11.20	45.56	12.50	36.71	8.65	42.03	11.44
Baseline 2	32.17	6.05	38.73	11.69	30.83	10.59	34.29	10.31
Condition Mean	37.46	5.29	42.14	9.21	33.77	8.06	-	-
Lane Deviation Mean (feet)								
Baseline 1	0.87	0.43	0.72	0.26	0.64	0.23	0.75	0.33
Baseline 2	0.86	0.46	0.71	0.18	0.64	0.26	0.74	0.34
Condition Mean	0.86	0.41	0.72	0.18	0.64	0.22	-	-
Lane Deviation STD (feet)								
Baseline 1	0.56	0.19	0.49	0.13	0.43	0.10	0.50	0.16
Baseline 2	0.57	0.20	0.49	0.11	0.43	0.07	0.50	0.15

Condition Mean	0.57	0.18	0.49	0.10	0.43	0.08	-	-

Table J.3 - Driving Performance Percent Difference Scores for the Conditions by Drive

Driving Block	No-Sound M	No-Sound SD	Spindex M	Spindex SD	Spearcon M	Spearcon SD	Drive Mean M	Drive Mean SD
Percent Difference Mean Follow Distance								
Drive 1	30.46	27.72	24.63	22.53	31.41	33.08	28.60	27.10
Drive 2	32.46	35.38	21.65	17.87	23.66	11.46	26.13	24.61
Drive 3	69.28	52.18	43.35	28.53	37.33	23.10	51.14	39.60
Condition Mean	44.07	29.69	29.87	18.96	30.80	16.38	-	-
Percent Difference STD Follow Distance								
Drive 1	35.65	42.63	42.78	47.62	52.45	40.32	43.38	43.39
Drive 2	51.82	39.87	26.30	15.05	45.29	23.37	40.35	29.25
Drive 3	198.42	131.10	153.70	72.26	164.38	76.69	171.49	96.14
Condition Mean	95.30	47.83	74.26	33.61	87.37	35.17	-	-
Percent Difference Mean Lane Deviation								
Drive 1	30.01	49.76	24.41	34.28	44.65	34.99	32.79	40.50
Drive 2	41.63	67.18	26.35	37.90	18.05	32.44	28.89	48.60
Drive 3	62.76	68.45	47.09	47.19	36.12	52.51	48.91	56.77
Condition Mean	44.80	55.44	32.62	30.12	32.94	28.81	-	-
Percent Difference SDLP								
Drive 1	39.31	38.23	43.75	44.54	60.57	47.63	47.88	43.70
Drive 2	52.33	59.20	27.58	32.68	31.28	29.45	37.06	43.10
Drive 3	66.76	61.19	50.48	30.44	52.29	47.72	56.51	47.66
Condition Mean	52.80	47.31	40.61	25.60	48.04	32.45	-	-

Table J.4 - Baseline Driving Visual Behavior Values for Conditions by Baseline Drive

Baseline Driving Block	No-Sound M	No-Sound SD	Spindex M	Spindex SD	Spearcon M	Spearcon SD	Baseline Mean M	Baseline Mean SD
Percent Time Off								
Baseline 1	0.60	0.50	1.35	1.69	1.10	1.62	1.06	1.44

Baseline 2	0.86	0.89	0.45	0.59	1.18	1.24	0.80	0.96
Condition Mean	0.73	0.50	0.90	1.00	1.14	1.21	-	-
Glance Count Rate (/min)								
Baseline 1	0.77	0.63	1.40	1.69	1.34	1.99	1.20	1.58
Baseline 2	1.04	1.11	0.53	0.74	1.13	1.20	0.87	1.03
Condition Mean	0.90	0.61	0.97	1.03	1.23	1.41	-	-
Mean Dwell Length (ms)								
Baseline 1	451.78	246.26	292.96	298.81	268.87	269.24	327.44	280.05
Baseline 2	534.82	82.90	260.43	347.41	466.72	361.05	407.80	322.21
Condition Mean	493.30	132.32	276.69	290.16	367.80	254.68	-	-

Table J.5 - Visual Behavior Difference Score Values for the Conditions by Drive

	No-Sound		Spindex		Spearcon		Drive Mean	
Driving Block	M	SD	M	SD	M	SD	M	SD
Percent Time Off								
Drive 1	31.98	6.92	21.75	10.23	25.43	14.58	26.47	11.68
Drive 2	26.78	6.03	9.69	6.23	16.14	9.67	17.68	10.22
Drive 3	23.31	5.79	5.93	3.38	14.30	9.69	14.70	9.79
Condition Mean	27.36	5.27	12.46	4.91	18.62	9.16	-	-
Glance Rate								
Drive 1	19.89	3.19	15.53	5.58	17.07	9.59	17.43	6.89
Drive 2	16.20	5.20	9.29	5.45	10.40	6.58	11.83	6.42
Drive 3	13.00	4.72	7.02	5.39	9.55	6.90	9.78	6.16
Condition Mean	16.36	3.66	10.61	4.56	12.34	6.97	-	-
Mean Dwell Length								
Drive 1	467.11	327.45	558.93	391.27	594.82	532.92	539.89	420.92
Drive 2	537.28	476.41	473.32	346.31	467.64	306.10	493.16	376.96
Drive 3	519.62	354.84	452.87	416.18	427.43	373.77	466.93	375.37
Condition Mean	508.00	346.13	495.04	331.78	496.63	276.74	-	-

Table J.6 - Heart Rate Measure Baseline Values for Conditions by Session

	No-Sound		Spindex		Spearcon		Session Mean	
Session	M	SD	M	SD	M	SD	M	SD
HR (BPM)								
Session 1	71.60	8.01	78.75	9.14	80.66	12.15	77.00	10.48
Session 2	69.79	7.99	78.67	12.38	85.79	14.09	78.08	13.28
Session 3	77.24	13.80	82.51	17.01	86.54	18.51	82.10	16.65
Session 4	74.67	10.03	82.74	14.08	79.85	9.73	79.09	11.70
Condition Mean	73.32	8.48	80.67	11.13	83.21	10.00	-	-

HRV								
Session 1	53.28	16.57	57.84	14.02	57.34	9.36	56.04	13.74
Session 2	53.36	13.65	64.00	12.20	61.97	12.20	59.55	13.35
Session 3	62.41	14.92	64.43	6.71	60.17	12.94	62.43	11.93
Session 4	59.09	15.09	63.97	15.77	57.64	13.60	60.32	14.83
Condition Mean	57.03	12.66	62.56	8.34	59.28	8.71	-	-

Table J.7 - Percent Difference Heart Rate Values for the Conditions by Drive

	No-Sound		Spindex		Spearcon		Drive Mean	
Driving Block	M	SD	M	SD	M	SD	M	SD
Percent Difference HR								
Drive 1	7.32	6.93	4.59	6.48	9.30	8.60	7.06	7.51
Drive 2	-1.15	6.93	-0.09	6.48	0.22	6.50	-0.32	6.51
Drive 3	-1.58	7.32	1.53	8.40	-0.39	5.48	-0.12	7.13
Condition Mean	1.53	5.42	2.01	5.98	3.04	5.51	-	-
Percent Difference HRV								
Drive 1	-10.87	20.32	-11.28	11.59	5.53	28.25	-5.52	22.42
Drive 2	-20.57	16.70	-12.90	12.25	-11.08	17.08	-15.21	15.92
Drive 3	-17.95	17.05	-6.86	22.83	-10.62	18.19	-12.34	19.32
Condition Mean	-16.47	14.19	-10.35	11.76	-5.39	12.76	-	-

Table J.8 - NASA-TLX Workload Scores for the Conditions by Drive

	No-Sound		Spindex		Spearcon		Drive Mean	
Driving Block	M	SD	M	SD	M	SD	M	SD
Mental Workload								
Drive 1	60.59	23.24	70.63	7.04	62.94	26.46	64.60	20.94
Drive 2	55.88	25.45	59.69	16.58	50.59	26.09	55.30	23.07
Drive 3	56.18	25.83	59.38	19.05	55.59	27.15	57.00	23.93
Condition Mean	57.55	21.67	63.23	12.51	56.37	20.78	-	-
Physical Workload								
Drive 1	27.65	23.26	28.75	14.89	41.47	22.69	32.70	21.29
Drive 2	36.76	24.43	32.50	18.07	47.06	24.56	38.90	23.02
Drive 3	35.59	23.31	33.75	15.65	50.88	25.08	40.20	22.77
Condition Mean	33.33	20.57	31.67	12.19	46.47	19.82	-	-
Performance								
Drive 1	65.00	17.03	57.67	19.07	40.77	18.80	55.34	20.47
Drive 2	50.00	22.21	35.00	6.55	26.54	9.44	37.95	17.53
Drive 3	47.50	19.75	38.67	11.09	31.15	13.72	39.66	16.54
Condition Mean	54.17	14.62	43.78	8.15	32.82	11.47	-	-
Effort								

Drive 1	65.59	23.38	75.67	8.42	60.88	24.51	67.04	20.99
Drive 2	52.35	23.59	62.00	16.67	51.47	24.42	55.00	22.08
Drive 3	55.00	24.17	67.00	12.65	54.71	25.40	58.57	22.10
Condition Mean	57.65	19.10	68.22	11.21	55.69	21.99	-	-
Temporal Workload								
Drive 1	55.29	18.91	65.29	19.72	44.41	15.09	55.00	19.65
Drive 2	52.35	23.06	46.76	24.11	39.12	22.86	46.08	23.52
Drive 3	55.88	25.81	50.29	20.42	43.82	22.40	50.00	23.07
Condition Mean	54.51	19.83	54.12	19.26	42.45	16.14	-	-
Frustration								
Drive 1	41.76	24.43	49.72	25.81	43.44	23.93	45.10	24.53
Drive 2	35.00	23.78	31.11	22.66	26.88	11.67	31.08	20.18
Drive 3	35.59	22.77	31.39	21.88	34.06	22.30	33.63	21.93
Condition Mean	37.45	20.15	37.41	20.68	34.79	15.35	-	-
Composite Workload								
Drive 1	60.04	18.53	66.67	8.69	55.00	19.86	60.57	16.84
Drive 2	52.57	18.16	52.96	14.30	48.27	21.18	51.27	17.86
Drive 3	53.43	19.79	56.59	12.94	51.55	23.62	53.86	19.02
Condition Mean	55.35	15.99	58.74	10.84	51.61	18.92	-	-

Table J.9 - Search Task Performance Data for the Conditions by Training Block

	No-Sound		Spindex		Spearcon		Block Mean	
Training Block	M	SD	M	SD	M	SD	M	SD
Mean Number Searches								
Block 1	15.88	3.46	15.94	5.24	13.13	4.21	15.06	4.51
Block 2	15.81	7.06	17.39	6.03	14.13	4.90	15.84	6.09
Block 3	17.94	6.80	17.61	6.82	13.69	5.26	16.49	6.52
Block 4	17.88	7.60	17.56	6.84	13.38	4.66	16.35	6.71
Block 5	19.00	7.42	18.72	6.68	14.44	5.14	17.47	6.71
Block 6	19.18	7.03	19.39	7.31	16.71	6.07	18.44	6.81
Block 7	16.81	6.34	18.00	7.34	15.65	5.93	16.84	6.52
Block 8	17.71	6.58	19.06	6.95	14.63	4.69	17.22	6.35
Block 9	17.00	7.75	19.39	6.99	15.88	6.04	17.46	6.99
Block 10	18.59	7.83	19.59	7.12	17.53	6.60	18.57	7.11
Block 11	19.18	7.86	20.56	7.29	17.88	5.70	19.23	6.96
Block 12	17.76	7.83	20.78	7.20	18.18	6.85	18.94	7.29
Block 13	18.53	8.57	19.94	6.61	19.75	5.27	19.41	6.86
Condition Mean	16.90	6.14	18.14	5.92	15.99	4.47	-	-
Percent Accuracy								
Block 1	91.06	5.82	86.82	7.69	87.55	13.79	88.42	9.75
Block 2	94.14	5.66	90.98	6.09	86.56	9.18	90.48	7.73
Block 3	95.10	4.83	94.45	5.46	89.87	12.42	93.10	8.53

Block 4	95.67	4.33	96.58	3.54	93.93	6.94	95.45	5.08
Block 5	96.72	3.59	95.78	4.78	88.75	9.88	93.71	7.49
Block 6	94.43	6.43	96.96	3.79	97.24	4.01	96.14	5.01
Block 7	92.06	7.67	94.61	6.43	95.18	6.37	93.90	6.86
Block 8	96.11	4.99	93.25	6.14	96.19	4.67	95.06	5.44
Block 9	91.62	8.73	96.37	4.30	91.01	8.83	93.00	7.81
Block 10	97.95	2.86	93.87	5.97	94.42	7.93	95.35	6.11
Block 11	95.04	5.43	97.44	3.95	95.33	6.01	95.98	5.17
Block 12	91.57	6.23	94.90	5.63	95.74	6.63	94.02	6.29
Block 13	95.86	3.51	96.87	3.64	98.25	3.04	96.99	3.48
Condition Mean	95.35	1.61	95.17	1.21	94.42	3.59	-	-
Mean Search Time (s)								
Block 1	24.66	5.58	25.45	8.06	27.55	9.71	25.87	7.89
Block 2	25.33	12.10	23.81	9.72	26.47	8.53	25.15	10.14
Block 3	22.80	10.25	23.33	9.06	27.32	10.46	24.40	9.91
Block 4	22.91	9.11	23.13	8.22	27.67	9.84	24.48	9.12
Block 5	21.11	8.93	22.69	9.50	25.94	10.22	23.17	9.55
Block 6	20.64	9.04	20.76	8.37	24.44	10.11	21.92	9.16
Block 7	22.85	9.43	21.56	8.68	28.73	13.87	24.47	11.21
Block 8	23.21	10.49	21.41	8.52	28.38	10.95	24.36	10.28
Block 9	23.53	9.46	22.05	9.14	23.98	7.97	23.12	8.75
Block 10	19.95	7.18	19.39	8.29	24.44	11.86	21.25	9.43
Block 11	19.74	7.87	18.37	7.63	21.93	7.63	20.02	7.68
Block 12	24.13	12.75	20.19	9.00	23.23	8.89	22.44	10.22
Block 13	22.99	11.78	20.75	8.21	19.96	6.47	21.22	8.96
Condition Mean	20.87	7.21	19.23	5.98	23.86	7.30	-	-

Table J.10 - Search Task Performance Data for the Conditions by Training Session

	No-Sound		Spindex		Spearcon		Session Mean	
Session	M	SD	M	SD	M	SD	M	SD
Mean Number Searches								
Session 1	15.88	3.46	15.94	5.24	13.13	4.21	15.06	4.51
Session 2	18.43	6.76	18.13	6.41	14.41	5.12	17.09	6.30
Session 3	18.16	7.23	19.45	6.99	15.75	5.23	17.90	6.63
Session 4	18.97	5.64	20.36	6.79	18.53	4.87	19.35	5.82
Condition Mean	17.86	5.77	18.47	6.36	15.46	4.86	-	-
Percent Accuracy								
Session 1	91.06	5.82	87.43	7.56	87.55	13.79	88.65	9.73
Session 2	94.97	2.67	95.41	1.15	91.28	6.89	93.83	4.69
Session 3	93.45	5.67	95.71	2.15	94.71	4.21	94.62	4.27
Session 4	96.31	3.94	94.76	4.52	95.70	3.71	95.59	4.02
Condition Mean	93.95	4.52	93.32	3.84	92.31	7.15	-	-
Mean Search Time (s)								

Session 1	24.66	5.58	25.45	8.06	27.55	9.71	25.87	7.89
Session 2	21.51	8.78	22.59	8.28	26.00	8.63	23.34	8.58
Session 3	22.96	10.64	20.63	8.18	23.67	7.77	22.34	8.82
Session 4	21.00	8.04	19.02	7.46	19.65	6.09	19.85	7.14
Condition Mean	22.53	8.26	21.92	7.99	24.22	8.05	-	-

Table J.11 - Ball Drop Performance Data for the Conditions by Training Block

	No-Sound		Spindex		Spearcon		Block Mean	
Training Block	M	SD	M	SD	M	SD	M	SD
Percent Accuracy								
Block 1	59.61	6.27	57.09	5.76	61.20	6.71	59.26	6.36
Block 2	58.69	7.07	56.81	5.80	60.33	7.20	58.58	6.73
Block 3	61.74	6.74	59.14	6.90	62.78	6.02	61.12	6.64
Block 4	63.07	2.89	60.38	8.34	62.06	4.82	61.71	6.06
Block 5	60.94	5.81	60.58	8.42	61.05	3.58	60.83	6.32
Block 6	61.43	8.89	62.61	8.36	63.35	7.83	62.46	8.24
Block 7	61.67	6.42	59.66	6.00	63.85	10.72	61.69	8.03
Block 8	62.68	8.66	60.64	7.40	63.87	11.18	62.36	9.10
Block 9	63.63	9.77	59.56	7.01	62.83	10.74	61.96	9.26
Block 10	62.30	7.01	59.63	7.03	62.45	10.60	61.39	8.35
Block 11	63.97	6.98	61.09	6.37	62.79	10.55	62.53	8.11
Block 12	62.14	5.12	64.28	5.76	65.50	7.88	64.05	6.43
Block 13	64.13	7.07	63.93	6.66	65.15	8.35	64.41	7.27
Condition Mean	61.83	2.86	61.23	5.13	63.15	4.60	-	-
Number Balls Dropped								
Block 1	481.29	66.87	467.06	58.09	476.59	65.09	474.98	62.47
Block 2	512.41	90.31	526.88	65.23	525.06	75.72	521.45	76.48
Block 3	540.18	93.66	549.29	61.78	548.47	91.67	545.98	82.07
Block 4	547.41	97.73	546.78	87.65	550.76	87.09	548.29	89.11
Block 5	545.41	103.22	570.00	65.02	544.00	85.61	552.80	85.49
Block 6	537.59	106.84	578.50	70.26	541.69	65.13	552.29	83.76
Block 7	576.82	103.80	570.44	89.78	584.71	89.90	577.19	92.94
Block 8	574.59	108.68	585.36	57.84	584.59	87.89	581.27	87.18
Block 9	582.41	111.71	582.93	74.31	575.41	84.29	580.14	90.28
Block 10	573.47	125.59	571.93	61.04	571.18	77.38	572.21	91.87
Block 11	585.24	118.26	584.44	70.10	574.94	79.80	581.48	90.41
Block 12	605.71	122.51	618.54	48.19	610.47	81.77	610.98	90.44
Block 13	606.35	125.40	626.87	64.54	604.06	70.62	611.84	90.67
Condition Mean	559.14	99.21	561.72	39.70	556.33	67.26	-	-

Table J.12 - Ball Drop Performance Data for the Conditions by Training Session

Session	No-Sound M	No-Sound SD	Spindex M	Spindex SD	Spearcon M	Spearcon SD	Session Mean M	Session Mean SD
Percent Accuracy								
Session 1	59.61	6.27	57.09	5.94	61.20	6.71	59.30	6.42
Session 2	61.18	6.57	60.32	7.10	62.11	6.57	61.21	6.65
Session 3	63.18	8.57	60.48	6.24	63.16	10.52	62.27	8.54
Session 4	63.11	6.51	64.10	5.88	65.33	7.77	64.18	6.69
Condition Mean	61.77	6.98	60.50	6.29	62.95	7.89	-	-
Number Balls Dropped								
Session 1	481.29	66.87	467.14	52.81	476.59	65.09	475.50	61.39
Session 2	536.60	95.22	543.10	62.33	543.80	78.25	541.05	79.10
Session 3	578.51	111.10	577.36	62.02	578.16	81.16	578.05	86.65
Session 4	606.03	121.66	619.61	50.97	607.27	73.34	610.43	87.32
Condition Mean	550.61	98.71	551.80	57.03	551.46	74.46	-	-

Table J.13 - Visual Behaviors Data for the Conditions by Training Block

Training Block	No-Sound M	No-Sound SD	Spindex M	Spindex SD	Spearcon M	Spearcon SD	Block Mean M	Block Mean SD
Percent Time Off								
Block 1	15.02	8.85	13.45	11.29	10.69	8.53	12.98	9.67
Block 2	13.03	11.87	10.66	10.32	10.21	7.16	11.33	9.89
Block 3	10.17	10.20	9.69	10.85	9.11	5.77	9.66	9.03
Block 4	11.94	13.30	9.41	10.43	5.99	4.43	9.18	10.25
Block 5	13.45	12.39	9.59	11.51	8.01	9.05	10.35	11.08
Block 6	11.51	11.55	9.26	10.89	6.24	5.07	9.06	9.73
Block 7	11.58	9.02	11.85	10.23	7.64	4.38	10.47	8.44
Block 8	9.30	8.13	11.06	12.04	7.89	4.44	9.48	8.82
Block 9	9.90	8.85	7.90	8.65	10.17	7.88	9.33	8.36
Block 10	11.22	10.28	12.84	13.38	10.10	6.33	11.39	10.25
Block 11	11.93	9.58	9.09	8.75	10.03	6.90	10.38	8.42
Block 12	8.06	8.45	7.77	6.20	11.89	8.36	9.32	7.81
Block 13	10.67	10.22	6.50	5.68	11.32	7.90	9.44	8.22
Condition Mean	9.67	7.76	8.42	7.80	8.62	4.79	-	-
Glance Rate (glance/min)								
Block 1	10.35	5.10	11.13	7.29	8.97	6.89	10.18	6.51
Block 2	9.03	7.19	7.98	5.71	7.75	5.43	8.24	6.04
Block 3	7.49	7.70	10.64	11.96	7.00	3.93	8.50	8.74
Block 4	4.82	4.17	6.36	5.53	5.08	3.20	5.46	4.39
Block 5	7.59	7.19	5.58	5.27	4.86	3.22	6.00	5.47

Block 6	6.07	5.71	6.34	6.11	5.07	3.50	5.83	5.17
Block 7	6.74	3.92	7.72	6.54	6.77	4.13	7.09	4.96
Block 8	6.06	4.18	7.70	7.70	7.55	5.67	7.10	5.96
Block 9	5.48	3.73	4.74	5.19	6.63	3.81	5.60	4.28
Block 10	5.67	4.08	7.23	7.39	7.11	3.64	6.67	5.24
Block 11	6.42	4.33	5.00	3.60	5.41	2.82	5.65	3.66
Block 12	5.35	3.83	8.03	7.42	5.77	3.16	6.47	5.34
Block 13	5.63	3.99	5.39	4.25	6.71	4.22	5.90	4.10
Condition Mean	5.94	3.42	5.89	4.66	5.66	3.11	-	-
Mean Dwell Length (ms)								
Block 1	781.08	193.06	601.06	221.87	628.47	262.97	669.35	237.46
Block 2	634.73	175.66	737.74	384.97	775.30	330.18	715.61	312.10
Block 3	776.08	380.24	828.77	490.26	684.45	226.29	764.20	380.72
Block 4	860.41	409.83	711.06	280.12	718.57	293.58	763.35	333.54
Block 5	923.89	309.06	903.44	552.64	693.35	314.61	841.10	421.70
Block 6	934.86	490.44	819.81	473.45	723.18	295.24	825.95	429.22
Block 7	782.64	360.52	694.35	208.94	689.37	199.25	722.12	264.11
Block 8	858.92	384.26	710.93	198.88	825.56	350.67	800.86	323.60
Block 9	922.91	449.90	807.98	352.93	860.34	488.66	862.44	428.38
Block 10	916.45	454.15	835.28	199.68	738.18	318.44	827.97	340.39
Block 11	1071.35	480.53	854.75	308.04	914.08	495.64	946.06	438.26
Block 12	816.09	370.15	718.43	191.03	767.88	323.23	766.44	298.24
Block 13	902.89	379.00	663.81	247.29	925.45	465.94	826.03	384.81
Condition Mean	812.53	271.18	689.79	157.45	694.97	203.64	-	-

Table J.14 - Visual Behaviors Data for the Conditions by Training Session

	No-Sound		Spindex		Spearcon		Session Mean	
Session	M	SD	M	SD	M	SD	M	SD
Percent Time Off								
Session 1	13.88	7.96	13.64	11.60	10.69	8.53	12.64	9.45
Session 2	9.03	8.65	9.72	10.52	9.65	8.50	9.49	9.08
Session 3	9.84	8.62	10.23	10.04	10.02	6.27	10.04	8.22
Session 4	8.76	8.72	7.16	5.86	11.58	7.85	9.23	7.60
Condition Mean	10.38	8.49	10.19	9.51	10.48	7.79	-	-
Glance Rate (glance/min)								
Session 1	10.35	5.10	11.18	7.51	9.49	6.80	10.36	6.49
Session 2	5.81	4.37	8.49	10.30	6.67	4.86	7.05	7.12
Session 3	5.50	3.19	6.83	6.25	6.57	3.72	6.33	4.59
Session 4	5.17	3.71	6.69	6.54	6.53	3.91	6.16	4.90
Condition Mean	6.71	4.09	8.30	7.65	7.31	4.82	-	-
Mean Dwell Length (ms)								

Session 1	781.08	193.06	589.08	224.23	628.47	262.97	667.03	239.41
Session 2	844.84	313.24	730.32	285.37	728.89	261.26	767.99	286.05
Session 3	919.54	390.29	788.64	271.59	836.33	366.17	849.16	345.17
Session 4	859.49	368.21	664.71	207.24	903.44	445.65	814.19	366.58
Condition Mean	851.24	316.20	693.19	247.11	774.28	334.02	-	-

Table J.15 - Heart Rate Percent Difference Data for the Conditions by Training Block

	No-Sound		Spindex		Spearcon		Block Mean	
Training Block	M	SD	M	SD	M	SD	M	SD
Percent Difference HR								
Block 1	4.12	6.42	4.56	8.10	3.64	7.39	4.11	7.18
Block 2	7.35	7.27	4.64	8.33	7.24	7.88	6.39	7.78
Block 3	6.09	8.23	4.02	10.82	2.90	4.84	4.33	8.32
Block 4	7.33	6.99	3.47	12.62	1.29	5.20	3.96	9.05
Block 5	7.53	7.46	1.09	8.40	1.47	6.56	3.36	7.92
Block 6	4.82	9.68	4.49	11.64	1.09	5.31	3.60	9.40
Block 7	3.18	6.29	11.04	13.31	4.43	4.04	6.19	9.34
Block 8	0.60	6.09	5.35	8.43	2.31	3.53	2.73	6.61
Block 9	-0.70	6.35	5.29	8.62	3.58	5.37	2.60	7.19
Block 10	-1.07	7.62	10.82	16.63	1.29	3.36	3.73	12.00
Block 11	-0.91	8.26	3.69	8.08	7.13	15.09	3.28	11.43
Block 12	5.08	5.64	6.73	6.36	3.31	2.79	5.23	5.43
Block 13	4.87	5.84	8.09	9.97	4.26	3.92	5.88	7.33
Condition Mean	3.41	5.57	4.56	6.18	2.38	2.34	-	-
Percent Difference HRV								
Block 1	-1.87	21.32	3.96	14.48	4.04	28.31	2.00	22.23
Block 2	-2.64	21.37	-1.19	24.59	-2.38	15.39	-2.06	20.48
Block 3	-7.08	29.18	-4.04	26.59	-0.70	15.61	-4.00	24.30
Block 4	7.25	24.45	-6.59	32.45	-6.65	21.29	-2.29	26.90
Block 5	-3.44	22.22	7.54	19.76	-5.51	14.00	-0.18	19.56
Block 6	-1.31	30.24	8.63	17.16	3.48	23.55	3.60	24.07
Block 7	-3.94	33.02	5.32	8.35	0.43	20.40	0.19	23.65
Block 8	-8.72	23.76	4.91	14.81	-1.73	16.76	-2.00	19.44
Block 9	-6.09	21.15	5.19	10.81	-9.14	12.48	-3.34	16.68
Block 10	-1.74	23.98	0.67	22.25	-1.06	20.90	-0.71	22.05
Block 11	-1.00	27.22	3.79	16.99	-0.64	22.86	0.62	22.58
Block 12	-2.83	19.29	2.11	22.63	9.08	19.42	2.49	20.69
Block 13	1.43	28.14	-1.33	22.40	1.55	27.52	0.55	25.62
Condition Mean	-4.24	17.72	1.09	9.44	2.92	3.85	-	-

Table J.16 - Heart Rate Percent Difference Data for the Conditions by Training Session

Session	No-Sound M	No-Sound SD	Spindex M	Spindex SD	Spearcon M	Spearcon SD	Session Mean M	Session Mean SD
Percent Difference HR								
Session 1	4.12	6.42	4.30	8.32	2.87	7.03	3.79	7.15
Session 2	6.32	7.87	4.51	9.27	3.15	5.25	4.73	7.64
Session 3	0.30	6.55	6.87	9.70	6.02	8.94	4.27	8.79
Session 4	4.75	5.53	7.21	6.75	3.63	2.82	5.22	5.43
Condition Mean	3.87	6.59	5.72	8.51	3.92	6.01	-	-
Percent Difference HRV								
Session 1	-1.87	21.32	3.57	15.06	0.32	28.32	0.47	22.07
Session 2	-2.48	22.25	1.60	18.10	0.67	16.55	-0.21	18.86
Session 3	-5.89	23.01	5.02	13.67	2.52	23.74	0.17	21.04
Session 4	-3.53	21.53	-5.13	17.54	5.68	26.28	-0.85	22.24
Condition Mean	-3.44	22.03	1.26	16.09	2.30	23.72	-	-

Table J.17 - NASA-TLX Workload Data for the Conditions by Training Block

Training Block	No-Sound M	No-Sound SD	Spindex M	Spindex SD	Spearcon M	Spearcon SD	Block Mean M	Block Mean SD
Mental Workload								
Block 1	82.00	9.78	83.33	12.77	74.41	22.97	79.68	16.69
Block 2	71.33	14.57	71.11	17.95	64.12	23.73	68.80	19.18
Block 3	76.07	9.64	71.47	13.20	71.25	20.12	72.77	14.96
Block 4	69.38	21.44	71.76	14.25	69.71	24.14	70.30	19.96
Block 5	77.31	7.80	70.88	11.76	70.29	22.67	72.45	15.88
Block 6	64.71	23.15	75.00	10.00	78.33	14.10	72.40	17.56
Block 7	60.88	23.27	67.78	14.06	69.38	18.61	65.98	18.92
Block 8	65.31	23.06	67.94	14.04	67.35	23.39	66.90	20.17
Block 9	62.06	23.85	68.53	14.77	73.44	16.71	67.90	19.09
Block 10	61.76	25.49	65.28	16.76	74.06	16.25	66.86	20.22
Block 11	60.88	26.82	63.06	20.08	75.94	16.04	66.37	22.09
Block 12	58.82	26.78	63.06	16.90	66.47	24.29	62.79	22.70
Block 13	60.29	27.64	63.06	17.83	67.94	24.24	63.75	23.24
Condition Mean	74.91	10.62	71.24	11.93	75.18	14.44	-	-
Physical Workload								
Block 1	48.24	26.51	49.17	32.51	65.00	25.50	54.04	28.92
Block 2	41.47	23.96	47.78	26.64	61.47	24.48	50.19	25.97
Block 3	43.24	24.55	51.94	24.74	63.53	27.32	52.88	26.39
Block 4	45.00	22.91	52.22	27.45	67.19	20.89	54.51	25.28
Block 5	44.41	25.36	55.00	26.57	65.29	28.03	54.90	27.50

Block 6	48.24	24.49	57.50	24.03	65.29	29.29	57.02	26.43
Block 7	40.59	26.51	48.06	22.76	59.41	26.39	49.33	25.92
Block 8	40.00	26.22	48.06	22.95	60.59	25.85	49.52	25.94
Block 9	42.94	24.56	48.33	24.55	62.06	26.52	51.06	26.00
Block 10	45.00	26.75	47.50	25.97	69.69	22.10	53.63	26.95
Block 11	43.82	25.71	45.56	26.34	69.06	22.97	52.35	27.14
Block 12	41.47	27.99	44.44	23.13	54.71	26.72	46.83	26.08
Block 13	42.06	28.56	47.50	23.34	59.71	29.55	49.71	27.68
Condition Mean	43.57	23.83	49.47	23.02	65.89	21.55	-	-
Performance								
Block 1	82.67	12.52	64.17	24.99	63.53	23.03	69.50	22.59
Block 2	71.47	13.20	64.41	15.80	65.29	21.90	67.06	17.30
Block 3	65.59	17.31	61.94	15.45	70.88	21.88	66.06	18.37
Block 4	73.67	10.08	58.89	15.77	70.88	22.72	67.40	18.08
Block 5	74.67	10.43	62.22	15.83	68.82	24.40	68.20	18.40
Block 6	76.67	12.91	60.00	17.15	64.41	23.84	66.50	19.59
Block 7	67.35	14.04	59.17	16.20	60.88	21.74	62.40	17.61
Block 8	66.76	14.57	60.00	15.72	60.88	25.01	62.50	18.83
Block 9	69.41	15.30	60.56	16.71	70.36	13.22	66.43	15.65
Block 10	74.00	12.56	60.56	15.14	77.50	13.55	69.89	15.55
Block 11	67.35	18.47	59.17	15.65	68.24	22.91	64.81	19.25
Block 12	68.13	16.92	56.67	15.81	58.53	23.70	60.88	19.38
Block 13	71.47	18.60	55.28	16.49	61.76	22.57	62.69	20.11
Condition Mean	74.62	9.00	59.41	12.81	69.73	12.85	-	-
Effort								
Block 1	86.00	9.30	77.35	13.12	82.50	11.11	81.77	11.69
Block 2	75.00	6.50	74.72	12.18	72.81	14.02	74.17	11.36
Block 3	67.19	12.64	70.28	10.77	78.33	12.91	71.73	12.69
Block 4	66.47	15.39	69.71	9.43	72.50	19.24	69.50	15.03
Block 5	61.76	22.08	72.22	10.60	80.36	11.00	70.92	17.07
Block 6	59.12	20.33	70.31	8.26	77.00	12.51	68.44	16.28
Block 7	66.47	18.60	62.65	10.77	66.56	18.23	65.20	16.00
Block 8	62.94	18.12	65.56	11.10	69.38	19.82	65.88	16.48
Block 9	61.76	19.68	63.89	12.19	74.67	12.46	66.40	15.91
Block 10	65.94	19.17	64.72	12.54	67.35	23.26	65.98	18.39
Block 11	60.59	19.03	64.72	13.98	74.00	13.12	66.10	16.30
Block 12	61.18	20.58	60.83	16.11	66.47	24.48	62.79	20.35
Block 13	60.59	20.98	62.22	15.07	69.12	24.95	63.94	20.56
Condition Mean	71.63	9.59	67.49	8.36	78.40	10.12	-	-
Temporal Workload								
Block 1	85.00	9.45	76.76	21.93	66.47	27.03	75.71	22.08
Block 2	72.50	14.14	80.31	11.03	62.06	21.80	71.43	17.80
Block 3	64.41	21.86	71.67	20.65	60.88	27.91	65.77	23.61
Block 4	65.59	20.45	70.00	21.07	63.24	24.68	66.35	21.88
Block 5	66.18	20.88	70.00	21.42	65.88	24.45	67.40	21.93
Block 6	67.94	21.87	70.56	22.02	65.59	24.80	68.08	22.56

Block 7	64.12	19.78	72.65	15.42	57.35	22.85	64.71	20.21
Block 8	66.18	21.47	68.89	19.97	67.19	20.08	67.45	20.13
Block 9	62.94	22.43	65.83	20.60	71.67	14.47	66.60	19.60
Block 10	61.18	26.31	65.00	23.83	62.35	24.82	62.88	24.54
Block 11	63.82	24.72	65.83	22.70	64.41	25.91	64.71	23.98
Block 12	62.65	26.58	66.94	23.34	59.12	23.40	62.98	24.20
Block 13	62.06	27.84	68.33	23.14	61.47	24.80	64.04	24.99
Condition Mean	70.95	16.10	75.36	14.55	69.21	14.84	-	-
Frustration								
Block 1	72.50	13.66	58.89	23.80	63.24	22.29	64.61	20.97
Block 2	57.06	21.00	52.50	19.42	58.53	28.27	55.96	22.84
Block 3	50.31	20.12	48.61	21.48	63.53	26.21	54.12	23.32
Block 4	54.71	23.55	52.22	22.18	63.82	26.61	56.83	24.19
Block 5	52.06	26.93	44.12	19.30	61.47	27.26	52.55	25.31
Block 6	52.06	25.19	47.50	23.96	61.47	31.66	53.56	27.19
Block 7	48.24	24.11	41.67	21.90	50.29	27.98	46.63	24.53
Block 8	32.31	11.48	43.06	19.11	56.18	29.08	44.79	23.34
Block 9	42.65	25.19	41.11	20.97	58.24	28.06	47.21	25.56
Block 10	42.33	20.52	39.44	19.62	57.65	27.85	46.50	23.95
Block 11	38.13	23.01	38.06	23.08	59.41	29.52	45.20	26.87
Block 12	38.67	20.57	34.44	21.00	50.29	28.37	41.10	24.17
Block 13	33.21	17.17	35.28	23.54	55.59	30.66	41.73	26.35
Condition Mean	40.71	8.71	42.60	16.71	58.44	26.20	-	-
Composite Workload								
Block 1	78.31	10.91	76.33	14.85	73.35	16.63	76.00	14.19
Block 2	70.27	11.20	71.63	11.97	70.35	15.75	70.78	12.79
Block 3	66.61	16.01	68.80	13.88	73.52	16.75	69.55	15.48
Block 4	67.49	15.95	68.35	11.97	70.22	21.58	68.68	16.58
Block 5	69.40	13.76	69.54	12.37	74.50	17.60	71.08	14.54
Block 6	68.67	14.26	69.44	11.70	75.56	17.52	71.15	14.60
Block 7	66.37	14.56	66.85	12.75	65.04	20.79	66.10	16.03
Block 8	65.45	16.20	66.56	11.09	69.69	17.24	67.17	14.77
Block 9	64.76	15.46	64.87	11.82	75.38	12.37	67.99	13.93
Block 10	65.47	18.47	64.22	14.39	76.09	11.01	68.21	15.66
Block 11	65.43	11.68	63.80	14.83	76.78	11.25	68.43	13.87
Block 12	63.82	18.42	63.13	15.10	64.29	22.14	63.74	18.34
Block 13	63.49	19.79	63.67	15.12	67.29	23.43	64.79	19.35
Condition Mean	70.28	11.31	67.49	11.83	74.70	12.87	-	-

Table J.18 - NASA-TLX Workload Data for the Conditions by Training Session

	No-Sound		Spindex		Spearcon		Session Mean	
Session	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Mental Workload								
Session 1	82.00	9.78	83.33	12.77	74.41	22.97	79.68	16.69

Session 2	71.74	15.41	71.62	12.76	68.90	22.43	70.67	17.28
Session 3	66.58	20.95	67.73	14.41	69.29	21.62	67.93	18.99
Session 4	64.33	25.08	64.17	16.74	67.21	24.11	65.32	21.93
Condition Mean	71.17	17.81	71.71	14.17	69.95	22.78	-	-
Physical Workload								
Session 1	48.24	26.51	49.17	32.51	65.00	25.50	54.04	28.92
Session 2	44.47	23.25	52.89	24.29	63.88	25.94	53.73	25.31
Session 3	42.47	24.97	47.50	23.60	62.65	25.70	50.81	25.73
Session 4	41.77	28.20	45.97	22.62	57.21	27.80	48.27	26.55
Condition Mean	44.24	25.73	48.88	25.76	62.18	26.23	-	-
Performance								
Session 1	82.67	12.52	64.17	24.99	63.53	23.03	69.50	22.59
Session 2	72.07	12.09	61.58	13.55	68.06	21.22	66.93	16.48
Session 3	68.95	14.16	59.89	14.26	65.22	21.16	64.42	16.97
Session 4	68.17	19.10	55.97	15.51	60.15	22.80	61.05	19.56
Condition Mean	72.96	14.47	60.40	17.08	64.24	22.05	-	-
Effort								
Session 1	86.00	9.30	77.35	13.12	84.00	9.67	82.23	11.36
Session 2	67.60	13.07	70.78	8.01	75.02	15.97	71.12	12.69
Session 3	67.33	15.03	65.07	11.24	70.82	17.59	67.63	14.58
Session 4	65.00	17.22	62.50	15.00	72.83	20.02	66.60	17.61
Condition Mean	71.48	13.66	68.93	11.84	75.67	15.81	-	-
Temporal Workload								
Session 1	85.00	9.45	76.76	21.93	66.47	27.03	75.71	22.08
Session 2	71.47	15.25	71.76	21.23	63.53	23.64	68.82	20.49
Session 3	68.67	18.33	67.51	21.09	62.54	23.47	66.14	20.91
Session 4	68.33	22.47	68.97	23.00	60.29	23.96	65.77	23.05
Condition Mean	73.37	16.38	71.25	21.81	63.21	24.52	-	-
Frustration								
Session 1	70.71	13.57	58.89	23.80	63.24	22.29	63.78	20.96
Session 2	50.57	15.92	49.40	18.63	61.76	27.06	54.03	21.64
Session 3	38.82	16.17	40.67	19.58	56.35	27.46	45.58	22.84
Session 4	38.57	20.73	34.86	21.73	52.94	28.98	42.19	25.07
Condition Mean	49.67	16.60	45.96	20.94	58.57	26.45	-	-
Composite Workload								
Session 1	78.31	10.91	76.33	14.85	73.35	16.63	76.00	14.19
Session 2	67.80	14.14	69.00	11.59	70.22	21.07	69.01	15.81
Session 3	65.41	15.00	65.52	12.64	68.23	20.56	66.39	16.13
Session 4	63.66	19.02	64.22	15.00	65.79	22.72	64.56	18.81
Condition Mean	68.80	14.77	68.77	13.52	69.40	20.24	-	-

Table J.19 - Survey Response Values for Conditions by Session

	No-Sound		Spindex		Spearcon		Session Mean	
Session	M	SD	M	SD	M	SD	M	SD

Effective at Search	3.18	1.07	3.44	0.86	3.35	1.17	3.33	1.02
Session 1	3.12	0.99	3.39	1.04	2.94	0.97	3.15	1.00
Session 2	3.53	0.94	4.44	0.86	3.65	1.27	3.88	1.10
Session 3	3.82	0.95	4.72	0.83	4.76	0.75	4.44	0.94
Session 4	3.41	0.99	4.00	0.89	3.68	1.04	3.70	1.01
Condition Mean	3.41	0.71	4.00	0.74	3.68	0.82	-	-
Display is Effective								
Session 1	3.76	1.44	3.89	1.08	3.41	1.33	3.69	1.28
Session 2	3.47	1.46	3.61	1.20	3.47	1.23	3.52	1.28
Session 3	3.41	1.28	4.06	1.00	3.71	1.40	3.73	1.24
Session 4	3.76	1.44	4.50	1.04	4.47	0.94	4.25	1.19
Condition Mean	3.60	1.26	4.02	0.79	3.77	1.05	-	-
Display is Helpful								
Session 1	3.59	1.50	3.83	1.20	3.53	1.01	3.65	1.24
Session 2	3.18	1.33	4.22	0.94	3.35	1.46	3.60	1.32
Session 3	3.29	1.31	4.00	1.03	3.65	1.22	3.65	1.20
Session 4	3.65	1.54	4.33	1.19	4.59	1.00	4.19	1.30
Condition Mean	3.43	1.31	4.10	0.88	3.78	1.00	-	-
Display is Annoying								
Session 1	3.06	1.44	3.22	1.22	4.06	1.44	3.44	1.41
Session 2	3.06	1.39	3.06	1.06	4.35	1.73	3.48	1.52
Session 3	2.82	1.55	3.06	1.06	4.12	1.62	3.33	1.51
Session 4	2.82	1.67	2.78	1.11	2.88	1.62	2.83	1.45
Condition Mean	2.94	1.38	3.03	0.96	3.85	1.39	-	-
Effective at Primary Task								
Session 1	2.76	0.97	3.06	0.80	3.18	0.81	3.00	0.86
Session 2	2.71	1.05	2.83	0.92	3.18	1.29	2.90	1.09
Session 3	2.82	1.07	3.22	1.00	2.88	1.05	2.98	1.04
Session 4	3.65	0.61	4.22	0.88	4.47	0.72	4.12	0.81
Condition Mean	2.99	0.73	3.33	0.58	3.43	0.69	-	-

APPENDIX K: TRAINING DATA BY BLOCK GRAPHS

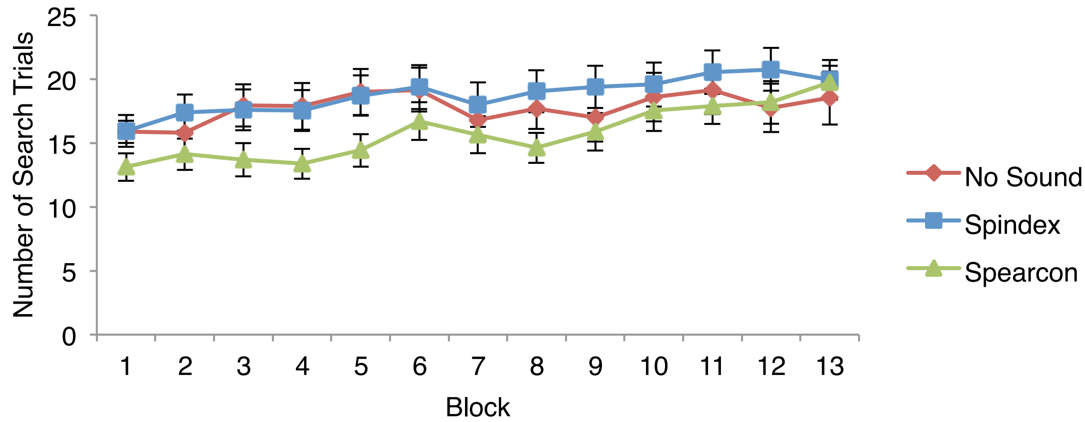


Figure K.1 - Graph of the mean number of search trials for each condition across the 13 training blocks (Session 1 = block 1; Session 2 = blocks 2-6; Session 3 = block 7-11; and Session 4 = blocks 12-13). Standard error is shown via error bars.

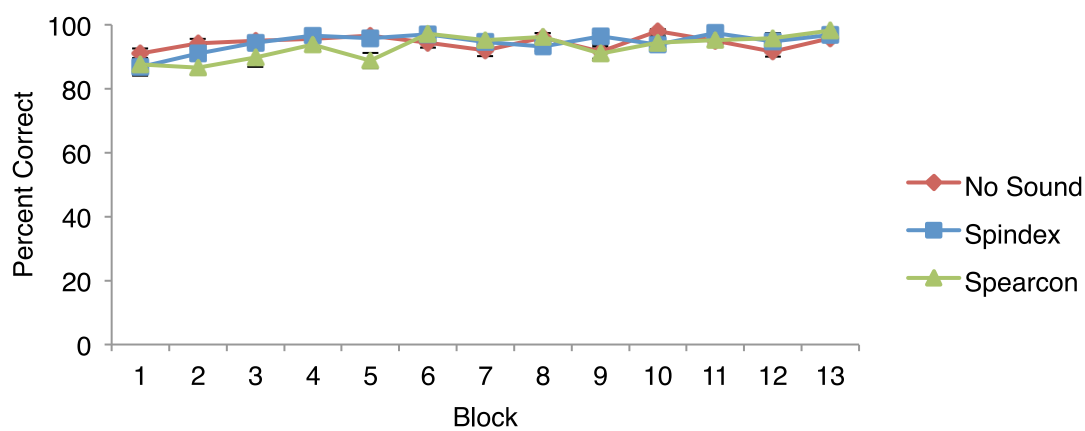


Figure K.2 - Graph of the mean percent correct song selections for each condition across the 13 training blocks (Session 1 = block 1; Session 2 = blocks 2-6; Session 3 = block 7-11; and Session 4 = blocks 12-13). Standard error is shown via error bars.

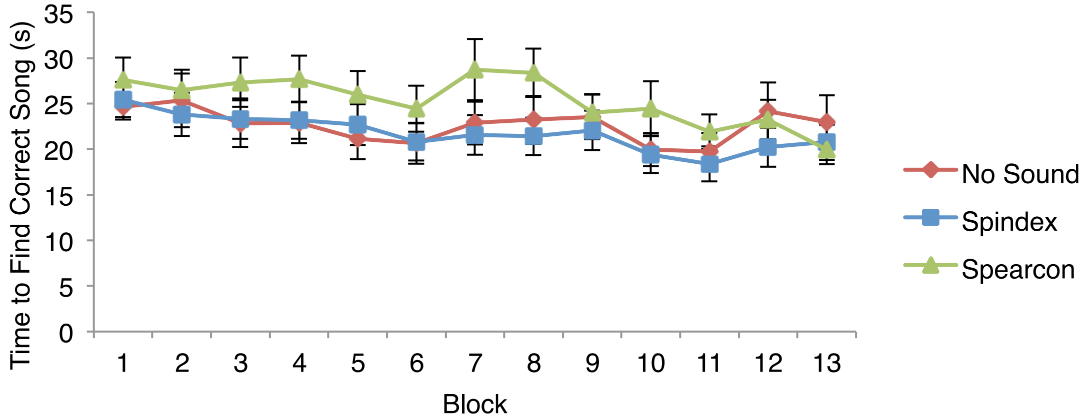


Figure K.3 - Graph of the mean time to find a correct song for each condition across the 13 training blocks (Session 1 = block 1; Session 2 = blocks 2-6; Session 3 = block 7-11; and Session 4 = blocks 12-13). Standard error is shown via error bars.

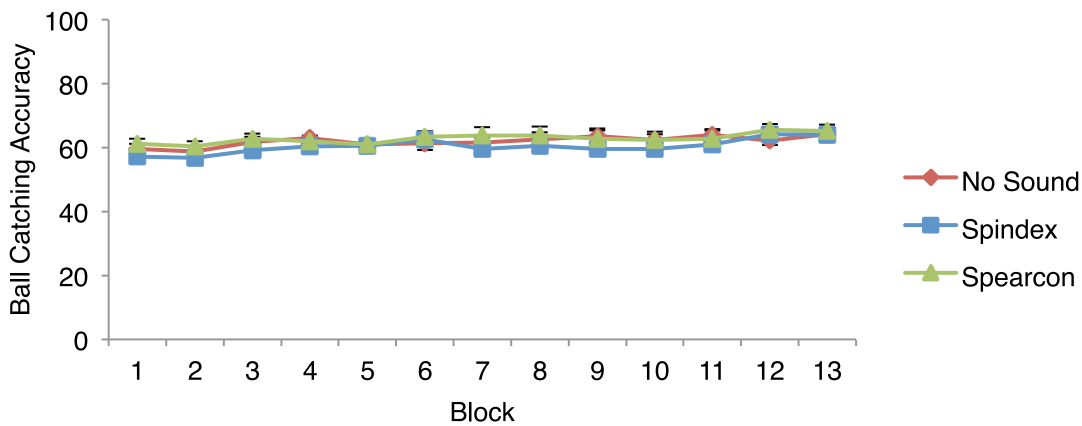


Figure K.4 - Graph of the mean ball catching accuracy for each condition across the 13 training blocks (Session 1 = block 1; Session 2 = blocks 2-6; Session 3 = block 7-11; and Session 4 = blocks 12-13). Standard error is shown via error bars.

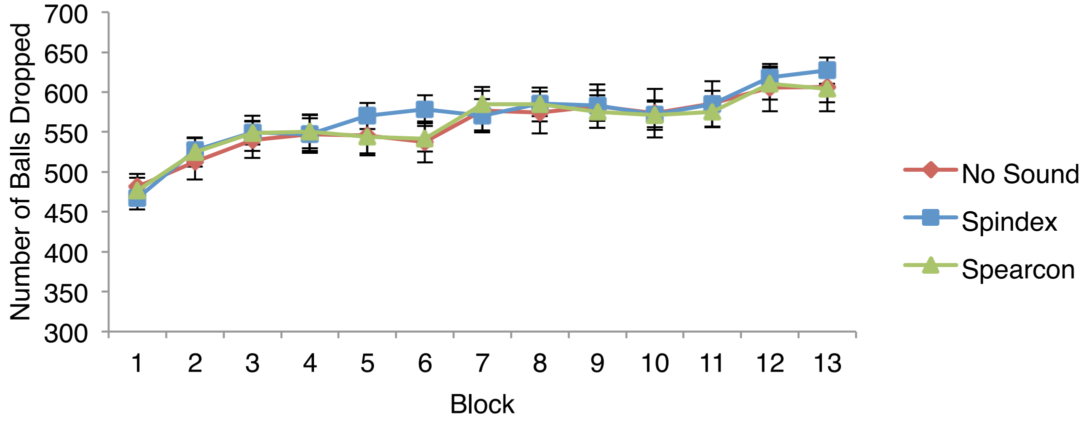


Figure K.5 - Graph of the mean number of balls dropped for each condition across the 13 training blocks (Session 1 = block 1; Session 2 = blocks 2-6; Session 3 = block 7-11; and Session 4 = blocks 12-13). Standard error is shown via error bars.

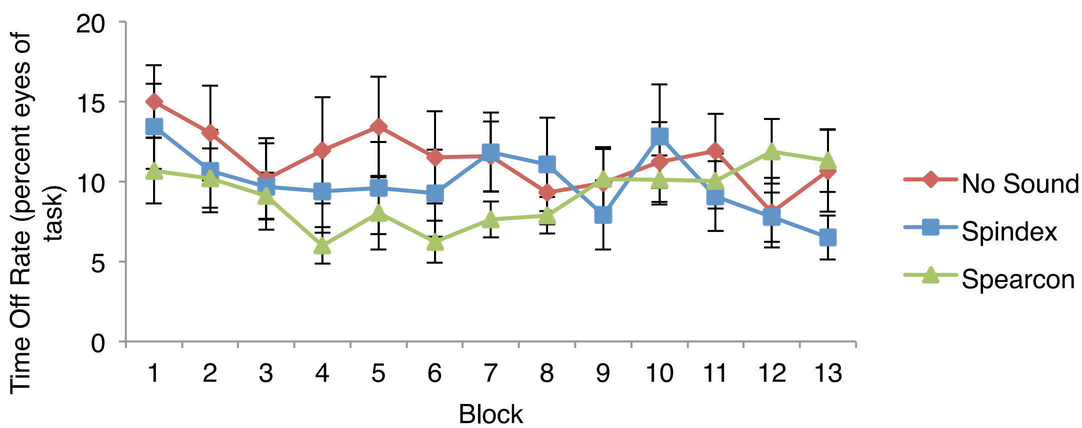


Figure K.6 - Graph of the mean time eyes off task rate for each condition across the 13 training blocks (Session 1 = block 1; Session 2 = blocks 2-6; Session 3 = block 7-11; and Session 4 = blocks 12-13). Standard error is shown via error bars.

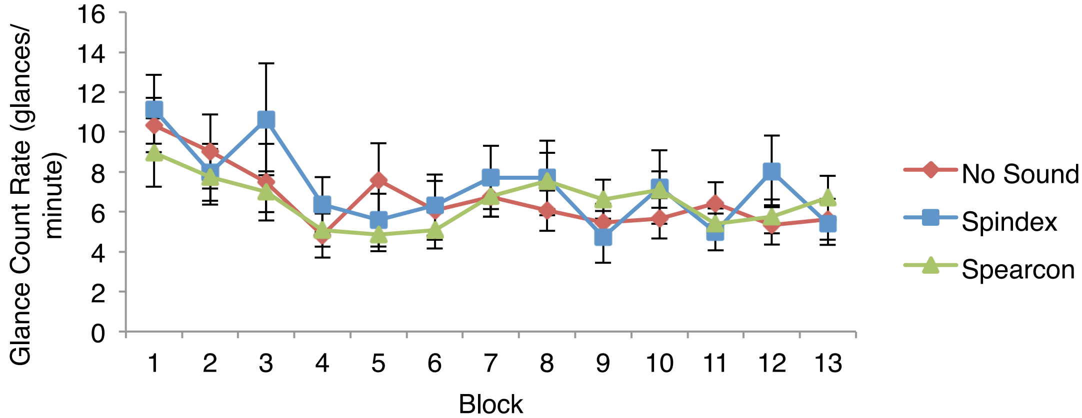


Figure K.7 - Graph of the mean glance count rate for each condition across the 13 training blocks (Session 1 = block 1; Session 2 = blocks 2-6; Session 3 = block 7-11; and Session 4 = blocks 12-13). Standard error is shown via error bars.

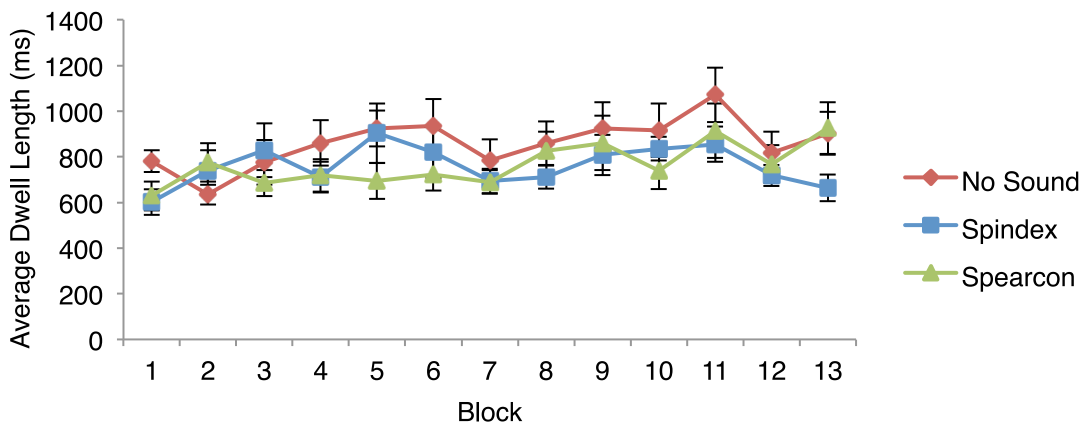


Figure K.8 - Graph of the average dwell length for each condition across the 13 training blocks (Session 1 = block 1; Session 2 = blocks 2-6; Session 3 = block 7-11; and Session 4 = blocks 12-13). Standard error is shown via error bars.

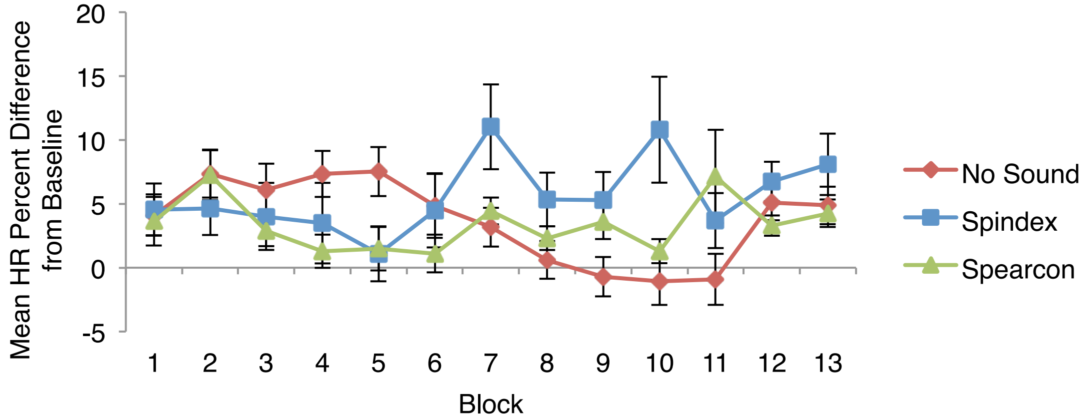


Figure K.9 - Graph of the mean percent difference of heart rate from baseline for each condition across the 13 training blocks (Session 1 = block 1; Session 2 = blocks 2-6; Session 3 = block 7-11; and Session 4 = blocks 12-13). Standard error is shown via error bars.

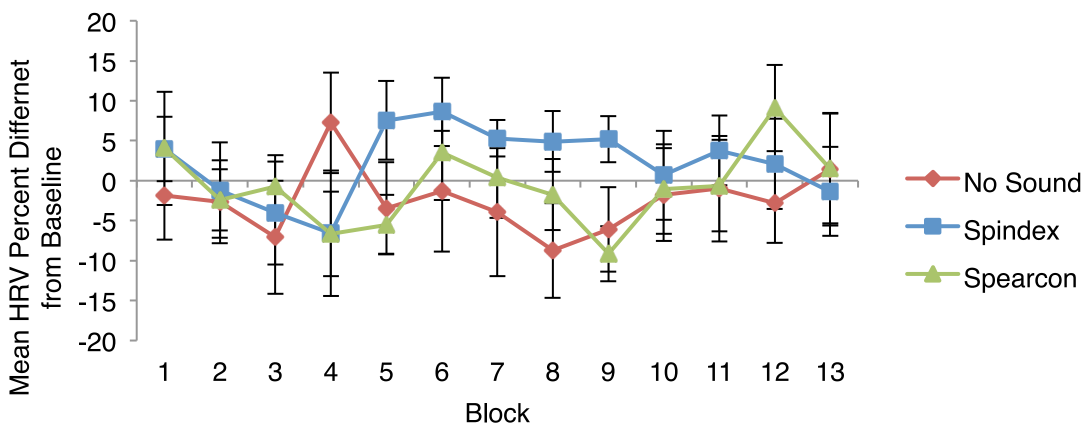


Figure K.10 - Graph of the mean percent difference of heart rate variance from baseline for each condition across the 13 training blocks (Session 1 = block 1; Session 2 = blocks 2-6; Session 3 = block 7-11; and Session 4 = blocks 12-13). Standard error is shown via error bars.

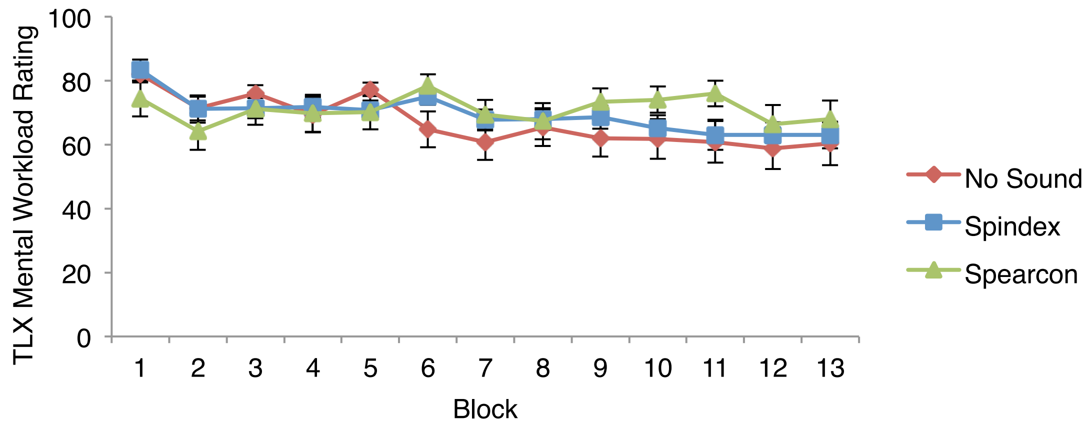


Figure K.11 - Graph of the mean subjective NASA-TLX mental workload rating for each condition across the 13 training blocks (Session 1 = block 1; Session 2 = blocks 2-6; Session 3 = block 7-11; and Session 4 = blocks 12-13). Standard error is shown via error bars.

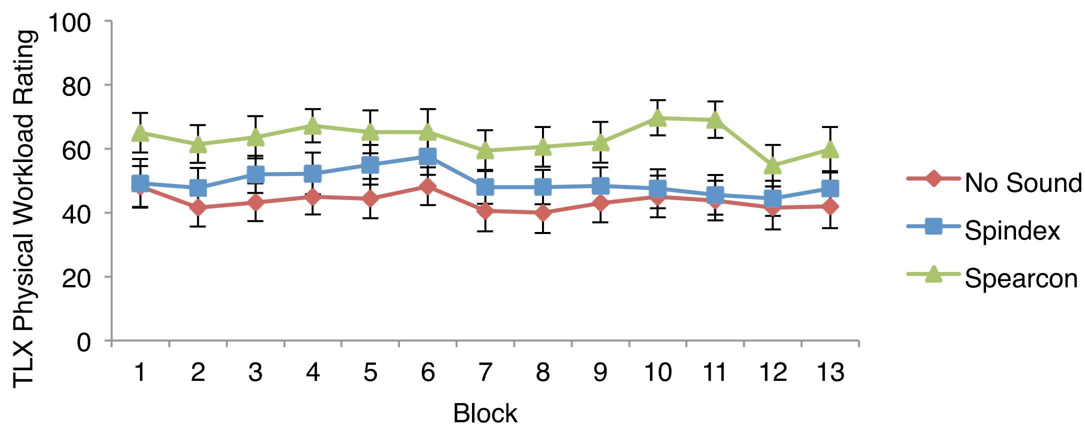


Figure K.12 - Graph of the mean subjective NASA-TLX physical workload rating for each condition across the 13 training blocks (Session 1 = block 1; Session 2 = blocks 2-6; Session 3 = block 7-11; and Session 4 = blocks 12-13). Standard error is shown via error bars.

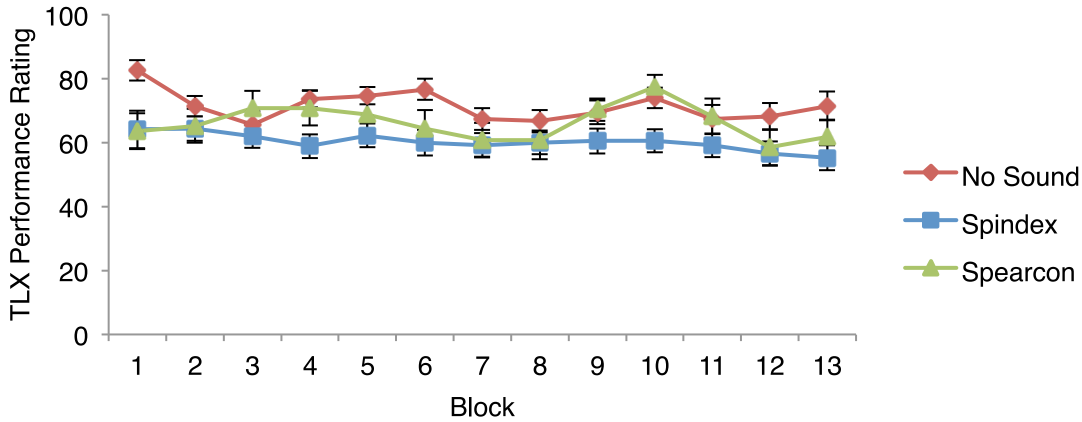


Figure K.13 - Graph of the mean subjective NASA-TLX performance rating for each condition across the 13 training blocks (Session 1 = block 1; Session 2 = blocks 2-6; Session 3 = block 7-11; and Session 4 = blocks 12-13). Standard error is shown via error bars.

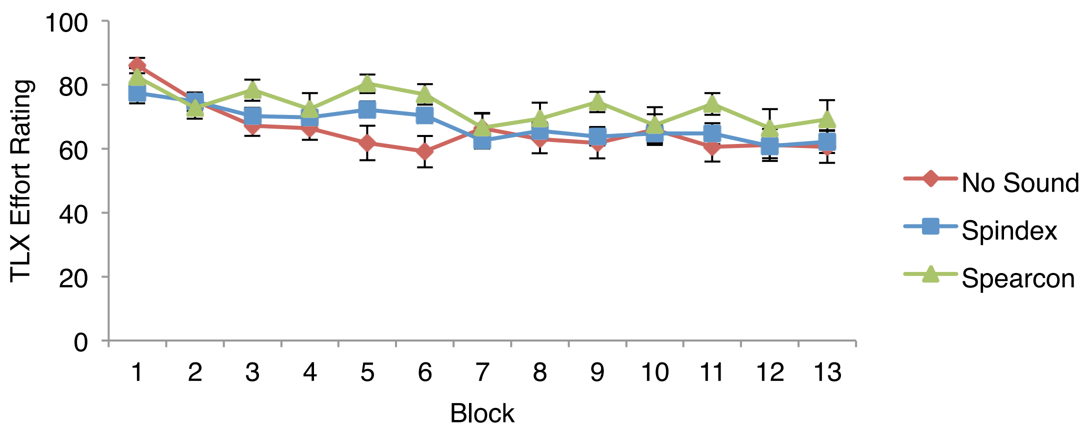


Figure K.14 - Graph of the mean subjective NASA-TLX effort rating for each condition across the 13 training blocks (Session 1 = block 1; Session 2 = blocks 2-6; Session 3 = block 7-11; and Session 4 = blocks 12-13). Standard error is shown via error bars.

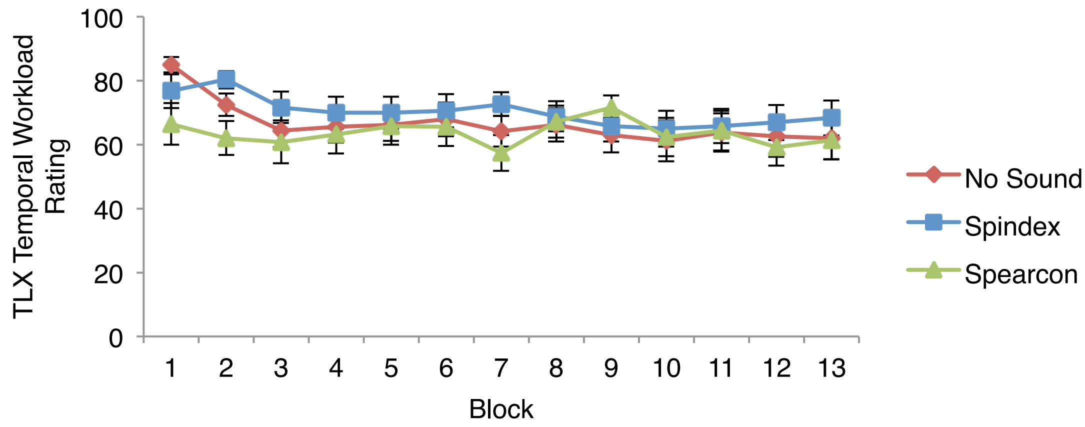


Figure K.15 - Graph of the mean subjective NASA-TLX temporal workload rating for each condition across the 13 training blocks (Session 1 = block 1; Session 2 = blocks 2-6; Session 3 = block 7-11; and Session 4 = blocks 12-13). Standard error is shown via error bars.

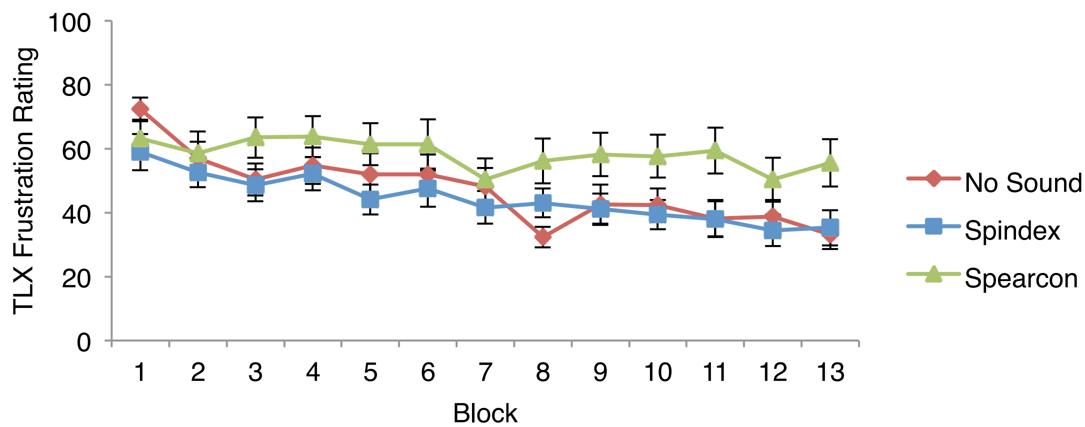


Figure K.16 - Graph of the mean subjective NASA-TLX frustration rating for each condition across the 13 training blocks (Session 1 = block 1; Session 2 = blocks 2-6; Session 3 = block 7-11; and Session 4 = blocks 12-13). Standard error is shown via error bars.

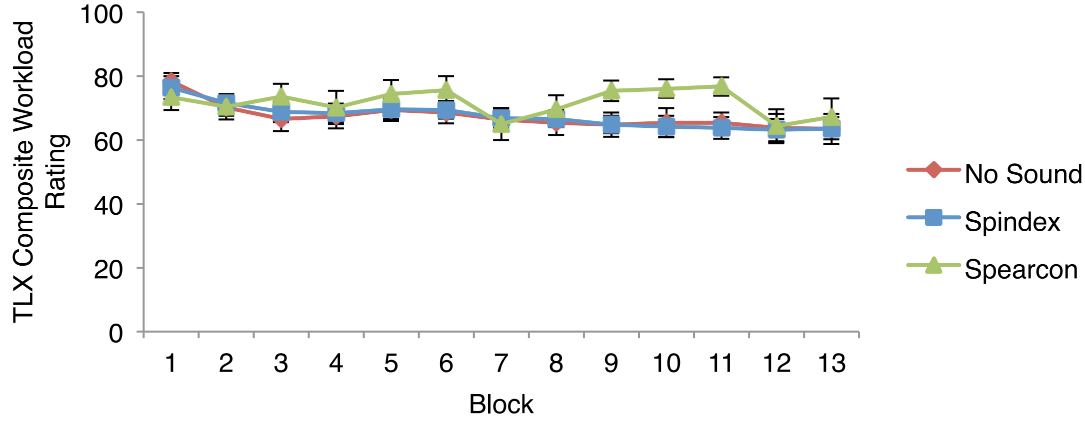


Figure K.17 - Graph of the mean subjective NASA-TLX composite workload rating for each condition across the 13 training blocks (Session 1 = block 1; Session 2 = blocks 2-6; Session 3 = block 7-11; and Session 4 = blocks 12-13). Standard error is shown via error bars.

REFERENCES

Brumby, D. P., Davies, S. C., Janssen, C. P., & Grace, J. J. (2011, May). Fast or safe?: how performance objectives determine modality output choices while interacting on the move. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 473-482). ACM.

Chisholm, S.L., Caird, J.K., & Lockhart, J. (2008). The effects of practice with mp3 players on driving performance. *Accident Analysis and Prevention*, *40*, 704-713.

Dingler, T., Lindsay, J., & Walker, B. N. (2008). Learnabiltiy of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech.

Drews, F. A., Yazdani, H., Godfrey, C. N., Cooper, J. M., & Strayer, D. L. (2009). Text messaging during simulated driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*.

Gable, T. M. (2015). Applying Spindex Auditory Cues While Driving and Performing A Secondary Search Task (Doctoral dissertation, Georgia Institute of Technology).

Gable, T. M., & Walker, B. N. (2013). Georgia Tech Simulator Sickness Screening Protocol. *Georgia Tech School of Psychology Tech Report GT-PSYC-TR-2013-01*.

Gable, T. M., Walker, B. N., Moses, H. R., & Chitloor, R. D. (2013, October). Advanced auditory cues on mobile phones help keep drivers' eyes on the road. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 66-73). ACM.

Hart, S. G. (2006, October). NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 50, No. 9, pp. 904- 908). Sage Publications.

Harvey, A. R., & Carden, R. L. (2009). Driving error and anxiety related to iPod mp3 player use in a simulated driving experience. *Perceptual and motor skills, 109*, 159-167.

Jeon, M., Davison, B. K., Nees, M. A., Wilson, J., & Walker, B. N. (2009, September). Enhanced auditory menu cues improve dual task performance and are preferred with in-vehicle technologies. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 91-98). ACM.

Jeon, M., Gable, T. M., Davison, B. K., Nees, M. A., Wilson, J., & Walker, B. N. (2015). Menu Navigation With In-Vehicle Technologies: Auditory Menu Cues Improve Dual Task Performance, Preference, and Workload. *International Journal of Human-Computer Interaction, 31*, 1-16.

Jeon, M., & Walker, B. N. (2011). Spindex (speech index) improves auditory menu acceptance and navigation performance. *ACM Transactions on Accessible Computing (TACCESS)*, *3*(3), 10.

Jeon, M., Walker, B. N., & Srivastava, A. (2012). "Spindex"(Speech Index) Enhances Menus on Touch Screen Devices with Tapping, Wheeling, and Flicking. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *19*(2), 14.

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J.

(2006). The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data.

Lamble, D., Kauranen, T., Laakso, M., & Summala, H. (1999). Cognitive load and detection thresholds in car following situations: safety implications for using mobile (cellular) telephones while driving. *Accident Analysis & Prevention*, *31*(6), 617-623.

Lasch, A. & Kujala, T. (2012, October). Designing browsing for in-car music player: effects of touch screen scrolling techniques, items per page and screen orientation on driver distraction. In *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 41-48). ACM.

Lewandowsky, S., & Thomas, J. L. (2009). Expertise: Acquisition, limitations, and control. *Reviews of human factors and ergonomics*, *5*(1), 140-165.

Liu, Y. C. (2001). Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveller information systems. *Ergonomics, 44*, 425-442.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological review*, *95*(4), 492.

Mehler, B., Reimer, B., & Wang, Y. (2011). A comparison of heart rate and heart rate variability indices in distinguishing single-task driving and driving under secondary cognitive workload. In *Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. 590-597).

National Highway Traffic Safety Administration. (2012). Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices. *Washington, DC: National*

Highway Traffic Safety Administration (NHTSA), Department of Transportation (DOT).

Nees, M. A., & Walker, B. N. (2011). Auditory displays for in-vehicle technologies. *Reviews of human factors and ergonomics*, *7*(1), 58-99.

Newell, A., Rosenbloom, P. S., & Anderson, J. R. (1981). Cognitive skills and their acquisition. *Mechanisms of skill acquisition and the law of practice*, 1-51.

Palladino, D. K., & Walker, B. N. (2007). Learning rates for auditory menus enhanced with spearcons versus earcons.

SAE (2013). J2944 Operational Definitions of Driving Performance Measures and Statistics (Draft). *Society of Automotive Engineers, Warrendale, Pa, USA.*

Schneider, W. (1985). Training high-performance skills: Fallacies and guidelines. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *27*(3), 285-300.

Sodnik, J., Dicke, C., Tomažič, S., & Billinghurst, M. (2008). A user study of auditory versus visual interfaces for use while driving. *International Journal of Human-Computer Studies, 66*, 318-332.

Speelman, C. P., & Kirsner, K. (1997). The specificity of skill acquisition and transfer. *Australian Journal of Psychology*, *49*(2), 91-100.

Stutts, J. C., Reinfurt, D. W., Staplin, L., & Rodgman, E. A. (2001). The role of driver distraction in traffic crashes.

Walker, B. N., Nance, A., & Lindsay, J. (2006). Spearcons: Speech-based earcons improve navigation performance in auditory menus. *In Proceedings of International Conference on Auditory Display, London, England* (pp. 63-68).

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, *3*(2), 159-177.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human factors*, *50*(3), 449-455.

Wilson, J. B. (2016). *Push and pull menus for auditory interfaces* (Doctoral dissertation, Georgia Institute of Technology).

Zhao, S., Brumby, D. P., Chignell, M., Salvucci, D., & Goyal, S. (2013). Shared input multimodal mobile interfaces: Interaction modality effects on menu selection in single-task and dual-task environments. *Interacting with Computers*, *25*(5), 386-403.